



PROCEEDINGS
OF THE
ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 201



LONDON

Published for the Royal Society by the
Cambridge University Press
Bentley House, N.W.1

23 May 1950

*Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)*

*and published by the Cambridge University Press
Cambridge, and Bentley House, London*

Agents for Canada, India, and Pakistan: Macmillan

CONTENTS

SERIES A VOLUME 201

No. 1064. 7 March 1950

	PAGE
Anniversary Address by Sir Robert Robinson, P.R.S.	1
The kinetics of the thermal decomposition of normal paraffin hydrocarbons. II. Comparative measurements on the series from propane to <i>n</i> -decane. By F. J. Stubbs and Sir Cyril Hinshelwood, F.R.S.	18
The slow oxidation of gaseous methyl ethyl ketone. By J. Bardwell and Sir Cyril Hinshelwood, F.R.S.	26
The contact between a solid and an electrolyte. By T. B. Grimley	40
The crystal structure of horse met-myoglobin. I. General features: the arrange- ment of the polypeptide chains. By J. C. Kendrew. (Plate 1)	62
The behaviour of supersonic flow past a body of revolution, far from the axis. By G. B. Whitham	89
Surface deformation and friction of metals at light loads. By J. R. Whitehead. (Plates 2 to 6)	109
Electronic wave functions. II. A calculation for the ground state of the beryllium atom. By S. F. Boys	125
The dispersion of a pressure pulse in the atmosphere. By R. S. Scorer . . .	137

No. 1065. 22 March 1950

The formation of a blast wave by a very intense explosion. I. Theoretical discussion. By Sir Geoffrey Taylor, F.R.S.	159
The formation of a blast wave by a very intense explosion. II. The atomic explosion of 1945. By Sir Geoffrey Taylor, F.R.S. (Plates 7 to 9)	175
On the reproduction and ruling of diffraction gratings. By Sir Thomas Merton, Treas.R.S.	187
The instability of liquid surfaces when accelerated in a direction perpendicular to their planes. I. By Sir Geoffrey Taylor, F.R.S.	192
The electronic structure of conjugated systems. VI. By B. H. Chirgwin and C. A. Coulson	196
The specific energy of crystal boundaries in tin. By K. T. Aust and B. Chalmers. (Plates 10 and 11)	210
Cellular atmospheric waves in the ionosphere and troposphere. By D. F. Martyn	216
Unified field theory in six dimensions. By J. Podolanski	234
One-dimensional dislocations. IV. Dynamics. By F. C. Frank and J. H. van der Merwe	261
Integration of the laminar boundary layer equation. I. Motion of an elliptic cylinder. Separation. By D. Meksyn	268
Integration of the laminar boundary layer equation. II. Retarded flow along a semi-infinite plane. By D. Meksyn	279
The interaction representation in the quantum theory of fields. By J. S. de Wet	284

	PAGE
The forces on a submerged spheroid moving in a circular path. By T. H. Havelock, F.R.S.	297
The thermal accommodation coefficients of gases. I. An investigation of the effect of flashing. By J. G. M. Bremner	305
The thermal accommodation coefficients of gases. II. Determinations at room and at liquid-oxygen temperatures. By J. G. M. Bremner	321
The collection of positive ions by a probe in an electrical discharge. By R. L. F. Boyd	329
The inelastic scattering of protons by magnesium, aluminium and other light elements. By E. H. Rhoderick	348
Paramagnetic resonance in copper sulphate. By D. M. S. Bagguley and J. H. E. Griffiths	366
The adsorption of vapours on mercury. IV. Surface potentials and chemisorption. By C. Kemball	377
The permeability of porous materials. By E. C. Childs and N. Collis-George	392
On the spontaneous magnetic field in a conducting liquid in turbulent motion. By G. K. Batchelor	405
World-structure and non-Euclidean honeycombs. By H. S. M. Coxeter and G. J. Whitrow	417

No. 1067. 23 May 1950

Wilkins Lecture. Robert Hooke. By E. N. da C. Andrade, F.R.S. (Plates 12 to 14)	439
Some theorems on perturbation theory. II. By E. C. Titchmarsh, F.R.S.	473
Studies on the ionization produced by metallic salts in flames. I. The determination of the collision frequency of electrons in coal-gas/air flames. By H. Belcher T. M. Sugden	480
The external gravitational and electromagnetic fields of rotating bodies. By G. L. Clark	488
Note on the problem of a rotating mass of perfect fluid in relativity mechanics. By G. L. Clark	510
Theoretical studies in nuclear structure. I. Enumeration and classification of the states arising from the filling of the nuclear <i>d</i> -shell. By H. A. Jahn	516
The electrical conductivity of thin wires. By R. B. Dingle	545
Spectroscopic studies of low-pressure flames. III. Effective rotational temperatures and excitation mechanism for C ₂ bands. By A. G. Gaydon and H. G. Wolfhard	561
Spectroscopic studies of low-pressure flames. IV. Measurements of light yield for C ₂ bands. By A. G. Gaydon and H. G. Wolfhard	570
Radially symmetric phase growth controlled by diffusion. By F. C. Frank	586
The far ultra-violet absorption spectra of the hydrides and deuterides of sulphur, selenium and tellurium and of the methyl derivatives of hydrogen sulphide. By W. C. Price, J. P. Teegan and A. D. Walsh. (Plates 15 and 16)	600
Index	610

Address of the President
Sir Robert Robinson, at the
Anniversary Meeting, 30 November 1949

Awards of Medals, 1949

The COPLEY MEDAL is awarded to Professor GEORGE CHARLES DE HEVESY, For.Mem.R.S., for his distinguished work on the chemistry of radioactive elements and especially for his use of isotopes as tracers in the study of biochemical problems.

Hevesy was one of the last to join the distinguished company of discoverers of elements in the classical tradition. In 1923, in collaboration with Coster, he established the occurrence of the element with atomic number 72 in zirconia minerals, and called it hafnium. This was shown to be a close analogue and constant comparison of zirconium, but a method of separation by chemical means was devised. The atomic weight was found to be 178.6, and the X-ray and optical spectra were fully described.

Here he wrote the last chapter of a volume and now we turn to the beginning of a new book.

When working under Rutherford in Manchester, Hevesy turned his failure to separate radium D from lead to good advantage. He recognized that the identity in chemical properties of radium D, and other radioactive isotopes of lead, with ordinary lead, made it possible to use these radioactive isotopes as indicators to follow the behaviour of lead in chemical processes, and in great detail owing to the extreme sensitivity of the methods of detection of radioactivity.

The first application of the idea was made with Paneth in 1913 at the Radium Institute of Vienna, and it was followed by a large number of interesting researches of a physico-chemical nature.

The use of radioactive isotopes as indicators or 'tracers' in biological processes was initiated in 1923 by Hevesy's studies on the uptake and distribution of lead in bean plants, using radioactive lead, radium D or thorium B, as indicators. After placing bean plants in solutions of ordinary lead nitrate containing small amounts of radium-D nitrate, the distribution of the lead was followed by determining the radioactivity of the ash from the different parts of the plants. This work may truly be said to have marked the opening of a new chapter in biochemistry. Furthermore, it established a pattern for the numerous subsequent researches by himself and by others which were to follow the discovery of artificial radioactive elements and the development of methods for the separation of certain stable isotopes.

Hevesy was one of the first to appreciate the potential biochemical importance of Urey's discovery of deuterium in 1932.

In the following year, by experiments with fish placed in water containing added D_2O , he and Hofer showed that there is a rapid exchange between environmental water and that in the body, and also that there is an exchange between the hydrogen

of the environmental water and labile hydrogen atoms in the tissue constituents. In 1934, by experiments on human subjects in which the density of the urinary water was compared with that of water ingested, they established the very important fact that at the low concentrations of D_2O present in ordinary water the body does not discriminate between D_2O and H_2O . It was found in the course of this work that the average time during which a water molecule remains in the body is 43 ± 1.5 days.

In 1937 Hevesy and his co-workers again broke fresh ground by their use of the radioactive isotope of phosphorus, ^{32}P , as a tracer in studies of the metabolism of phosphorus compounds. For example, ^{32}P -labelled phosphate was administered to rabbits, and its uptake by various tissues and its rate of excretion were determined. It was found that the average time during which a phosphorus atom remains in the body is thirty days. Having established by *in vitro* experiments that there is no direct exchange between organic ester phosphate and inorganic phosphate, Hevesy and his co-workers were able to determine the 'turn-over' rates for certain organic phosphorus compounds in the body by isolating these compounds at intervals after the administration of labelled phosphate and estimating their content of radioactive ^{32}P .

Work on the uptake of labelled phosphate by bone did much to establish that the whole of the skeleton is in a state of dynamic equilibrium with the constituents of the body fluids. By studies on lactating animals Hevesy and his collaborators were able to show that ^{32}P is incorporated into the caseinogen of milk within a few hours. They obtained ^{32}P -labelled adenosine triphosphate enzymatically from labelled inorganic phosphate, and clarified the role of this compound as a phosphate donor in carbohydrate metabolism. Further, their studies on the phosphatides in the liver and blood plasma supported the view that the plasma phosphatides are synthesized in the liver.

The spark of the Manchester days has kindled a flame that can illumine all that is dynamic in biochemistry.

A ROYAL MEDAL is awarded to Sir GEORGE PAGET THOMSON for his distinguished contributions to atomic physics and especially for his investigation of the wave properties of the electron.

During a period of more than thirty years, Thomson has made many notable contributions to experimental and theoretical physics over a wide range of subjects.

In 1914-18 he developed important aspects of the science of aeronautics and took part in experimental flights in order to investigate problems of interest to the air services. His results were communicated in Government reports, and at the end of the war he wrote a book *Applied Aeronautics*.

Early work in the laboratory was, with filial piety, connected in some way or another with positive rays, and in this field he disclosed a number of important results.

But Thomson's most distinguished contribution to knowledge is his demonstration of the wave nature of the electron. Shortly after, but independently of Davisson's discovery of the diffraction of electrons by single crystals, Thomson found that

a narrow pencil of swift electrons, after transmission through a thin film of polycrystalline matter, produced, on a photographic plate or fluorescent screen, a pattern of rings analogous to optical halos or to the Debye-Scherrer rings well known in the corresponding experiment with X-rays. By simple and beautiful experiments he was able to show that the patterns could only be explained on de Broglie's revolutionary theory of the wave nature of matter, and he proved, within the experimental accuracy of 1%, that the wave-length associated with the electron must be h/mv , as required by the theory.

In addition to bringing conclusive proof of this remarkable duality in the behaviour of matter, these experiments opened out a new and fruitful field of research which has important practical applications. Thomson showed that electron diffraction provided an eminently suitable method for the study of the structure of surfaces. As a result it is now possible to investigate how the structure of the surfaces of metals is changed by mechanical, thermal or chemical treatment, and thus to extend by a powerful method the information about the structure of matter which can be obtained by the complementary method of X-ray diffraction.

In later years Thomson's interests have extended to nuclear physics and important contributions on some aspects of this subject, including the effects of cosmic rays and the properties of mesons, have come from his laboratory.

Thomson's work has been in the direct line of progress; it has given us a new conception and a new tool for research. In both respects it has notably widened our horizons.

A ROYAL MEDAL is awarded to Professor RUDOLPH ALBERT PETERS for his distinguished researches in biochemistry and in particular his discovery of the role of vitamin B₁ in tissue metabolism.

The dramatic recovery of an animal suffering from some particular vitamin deficiency when given a minute amount of that vitamin, coupled with the observation that the curative dose must be repeated at intervals to prevent a relapse to avitaminosis, had drawn attention to the possibility that vitamins were either catalysts or played some important role in essential enzyme system.

But it was the work of Peters on the brain tissue of vitamin B₁-deficient pigeons, begun in 1929, which showed for the first time that the latter suggestion was indeed correct. In 1931 he found that polyneuritic symptoms in such pigeons could be correlated with a general deficit in lactic acid metabolism and that this was most marked in the lower regions of the brain. Comparing normal and avitaminous brain tissue he concluded that B₁ must play some essential part in tissue respiration, namely in the glucose-lactic acid system. Following the elucidation by R. R. Williams and others of the chemical structure of B₁ (aneurin) he was able to show that the vitamin, in its phosphorylated form as co-carboxylase, was part of the pyruvic oxidase system and was thus an essential factor in carbohydrate metabolism.

In 1937 Peters found that pyruvic oxidase, on which he was working in connexion with vitamin B₁, was poisoned by small amounts of arsenite; later, with Sinclair and Thompson, he showed that it was sensitive also to certain organic arsenicals. This

led him to suggest that the toxicity of such substances was due to their action on the functional thiol groups of this enzyme.

It transpired that the more potent vesicants, containing the group AsCl_2 , combined with two thiol groups of kerateine (Stocken & Thompson), and this suggested the use of a synthetic dithiol such that an arsenic atom, two sulphur atoms, and two or three carbon atoms might be locked together in a stable ring system. The object would be to protect the enzyme by exhibiting, in effective mass, an alternative anchor for trivalent arsenic. A new dithiol, now known as B.A.L., was selected as the most promising representative of its type, and this was found to shield the pyruvic-oxidase system from arsenicals and even to be capable of reversing a toxic action when introduced at a later stage. B.A.L. has been found to be effective against arsenical vesicants in man and to show a beneficial action against arsenical dermatitis in a substantial number of cases. It has recently been found to be useful also in acute mercurial poisoning and in the dermatitis that may follow administration of compounds of gold.

There can be little doubt that the discovery of B.A.L. by Peters and his associates represents an advance of considerable importance in practical therapeutics. It was clearly no chance observation but the result of a logical development conforming to the valid principles of scientific method.

We can make for Peters the proud claim that he was the first to elucidate the role of a vitamin in animal metabolism.

The DAVY MEDAL is awarded to Professor ALEXANDER ROBERTUS TODD for his analytic and synthetic studies in organic chemistry and biochemistry with special reference to vitamins and nucleosides.

In early collaborative work on the bile acids with Borsche the Wieland-Windaus constitutions current at that time were shown to be inadequate. Later, at Oxford, Todd made a very substantial contribution to the synthesis of diglycosidic anthocyanins, and he also helped to determine the structure and to effect the syntheses of mould products of the anthraquinone series, such as helminthosporin and cynodontin. He also established the main lines of the constitutions of flavoglauцин and auroglauцин.

His work on aneurin was of great value: he took a part in the determination of the structure of vitamin B_1 and developed syntheses, one of which is now employed for the manufacture of the vitamin. In the course of this investigation he established the structure of thiochrome and proved it by synthesis. He determined, in parallel with Fernholz & Karrer, the structure of vitamin E and effected its synthesis.

His studies of cannabinol, the active principle of hashish, were also crowned by synthesis.

From this stage on, his work has been characterized by increasing and individually characteristic originality. He has provided the last details of the structure of all the natural purine and pyrimidine nucleosides, and has synthesized them. A resounding success was the synthesis of A.D.P. and A.T.P., or adenosine triphosphate, and the artificial preparation of cozymase has been brought near to its final stages. This work involved much attention to pyrimidine chemistry, to some aspects of the

carbohydrates, to the processes of phosphorylation, and to the properties of phosphoric esters and anhydrides. It is of quite outstanding merit, a milestone in the progress of biochemistry.

Other problems which Todd has illuminated, and often solved, concern the nature of the specific germinating factor for *Striga hermonthica* (this turned out to be D-xyloketose); the nature of the factor produced by solanaceous plants and which is required for the hatching of eelworms; the mechanism of the hardening process of insect cuticle. He has recently cleared up the chemistry of a series of *bis*-isoquinoline alkaloids and started an intensive study of insect pigments, especially the remarkable colouring matters of Aphididae, which exhibit such interesting colour changes.

Todd has proved himself a master of the strategy and tactics of research, and the great school which he directs at Cambridge has enhanced the prestige of British chemistry throughout the world. His achievements have already added greatly to our knowledge of organic chemistry and biochemistry, and still more they have indicated sure pathways for further advance towards the solution of some of the central problems of biology. Can it be an exaggeration to describe in this way work which is essential for a better understanding of at least one aspect of the fundamental chemistry of nucleo-proteins?

The SYLVESTER MEDAL is awarded to Professor LOUIS JOEL MORDELL for his distinguished researches in pure mathematics, especially for discoveries in the theory of numbers.

Although he is also a powerful analyst, Mordell has always been primarily an arithmetician. Indeed, he was for long almost the only British mathematician of whom this could be said, and, if this is no longer true, it is mainly the result of his own teaching and example.

His most enduring interest has been in the theory of indeterminate equations. This is a subject which has attracted many great mathematicians, but it is one of supreme difficulty and new ideas emerge only at rare intervals. One of Mordell's greatest achievements was his proof in 1921 of the 'finite basis' theorem for the rational solutions of a cubic equation $f(x, y) = 0$ in two variables. This asserts, roughly speaking, that all the rational solutions can be derived, by a systematic process, from a finite number of them. The theorem represented a great advance in the theory of indeterminate equations, and has inspired a great deal of work by other mathematicians all over the world. In other papers, from his earliest to his most recent, Mordell has developed a variety of methods which throw new light on many different types of indeterminate equations.

Another field in which Mordell has made great advances is that of the geometry of numbers. One general problem here is that of finding the best possible inequality for the minimum of an algebraic form of a given type, when the variables take integral values. The results for a binary quadratic form are classical. The binary cubic form was considered by several mathematicians of the last century, but the final result was found by Mordell as late as 1940. The geometrical interpretation of the problem involves a particular non-convex region in the plane, and the methods which Mordell devised for its solution proved to be applicable to other non-convex

regions. They enabled problems to be solved which would formerly have been considered unapproachable.

A third region in which Mordell has shown his mastery is the theory of the elliptic modular functions and their applications to the theory of numbers. In particular, he used the theory to give a new treatment of the representation of numbers as sums of squares, and to prove the validity of some of Ramanujan's conjectures.

Mordell's output relating to these topics and to numerous individual problems is characterized by the subtlety and fertility of the methods he has discovered. He has had a profound influence on the development of the theory of numbers, and has been widely recognized as one of the most eminent mathematicians of our time, both for the importance of his own researches and for his inspiration of the work of others.

The HUGHES MEDAL is awarded to Professor CECIL FRANK POWELL for his distinguished work in experimental physics and especially for the discovery of mesons and their transformations.

Powell's early researches were concerned with the properties of ions. These led to the investigations concerning fundamental particles and atomic nuclei using the photographic plate technique, for which he and his school of experimental physics at Bristol have a world-wide reputation.

He was instrumental in bringing about a marked improvement in the quality of the sensitive material especially developed for research in nuclear physics, and using these improved plates, has discovered a new fundamental particle, the π meson, of which the mass is 280 times that of the electron. The negative π mesons are captured by atomic nuclei, causing the nuclei to disintegrate. The positive π mesons decay spontaneously, when at rest, into μ mesons of 205 electron masses, whose existence had been established by other means, and these in turn into a positive electron and presumably a neutrino.

These discoveries relating to mesons have proved to be of profound importance in unravelling the complicated phenomena of cosmic radiation in the earth's atmosphere and have thrown new light on the theory of nuclear forces.

High-energy nuclear disintegration can so far only be studied by means of cosmic rays, and Powell has also been responsible for several remarkable discoveries in this field.

All these and other nuclear transformations have been recorded in a most striking and beautiful way in the photographs of Powell and his school.

In him we salute a master of advanced technique which has made possible discoveries that are certainly of transcendent natural philosophical importance. They may well prove to be also of great practical significance in unsuspected directions in relation to the physical system as a whole.

I wish to thank Members of Council for the great help they have given me in the preparation of notes on the Medallists.

I do not propose on this occasion to expatiate on the matters raised in the Report of Council or to allude to any of the more general topics, which concern our interest.

The reason is that several important developments are in process, and a statement made at the present stage would have little value.

As the result of the ballot has been announced I am entitled to hope that the position in regard to the most important of current activities, namely, accommodation of the scientific societies, will be clearer in a year's time, and that I shall be able to announce definite progress.

There is continuous, though slow, improvement in the speed of publication of papers in the Society's *Proceedings* and *Transactions*. An important factor is the time taken by referees, and the majority of Fellows have responded with corresponding willing action to the request that referee forms should be returned within three weeks of the receipt of a paper.

The Secretaries and Publication Committee hope for still further acceleration of this stage of the work.

On behalf of the Fellows of the Society I take this opportunity of expressing to the British Thomson-Houston Company our warm thanks for their generous action in financing the installation of fluorescent lighting in our rooms.

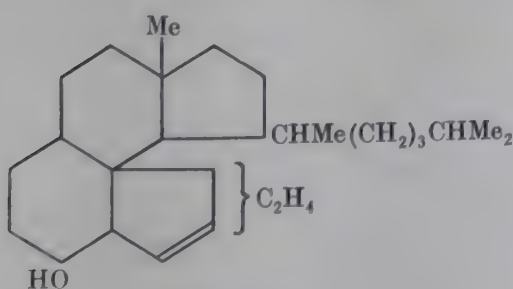
THE SIGNIFICANCE OF THE STERINOIDS

There are certain sections of organic chemistry of such a specialized nature that only a few persons have a real mastery of their intricacies. Examples are the chemistry and technology of azo-dyes, or of anthraquinone derivatives and related polycyclics. And certainly the chemistry, biochemistry and physiology of the sterinoids cannot be comprehended fully except by those who devote undivided attention to these subjects. It would clearly be impossible to describe any part of this vast field at all adequately in the few minutes at my disposal, and I propose to do no more than glance at a few of the headings of developments in the last twenty years. My object is to justify a plea for an even greater effort in research and for international co-operation such as was achieved during the war in the penicillin field.

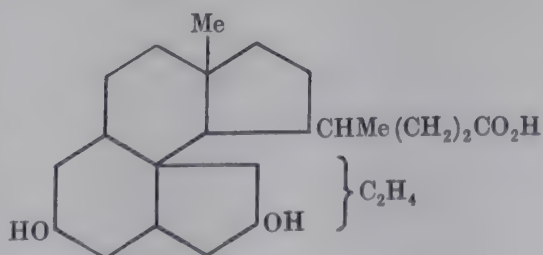
Although cholesterol is a constituent of all animal cells there was no inkling of its physiological importance when Windaus began his investigations in 1903, and this was also true in 1912, the approximate date of commencement of Wieland's work on the related bile acids.

The molecular structures were studied as interesting organic chemical problems, curiosities, attractive because of their difficulty and the unique character of the group. Nevertheless, the impressive accumulation of facts in this period established many of the fundamental relations and laid the foundation for further progress.

The formulae that were regarded as correct for several years were:



cholesterol (1928)

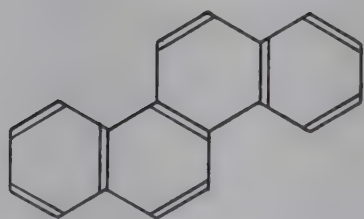


desoxycholic acid (1928)

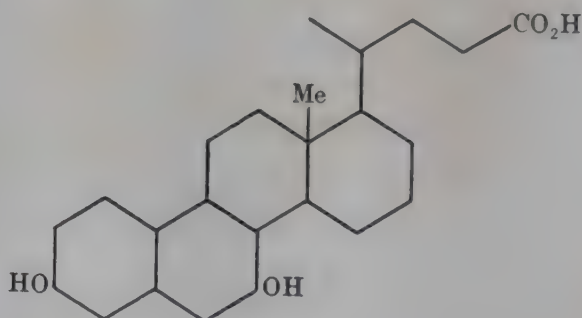
Various attempts to accommodate the group C_2H_4 in these structures were unsuccessful, so that it may be said that no full and satisfactory constitution was advanced. It was gradually realized that the problem was quite unsolved in regard to the precise nature of the polycyclic portion of the molecule, though the evidence for the side-chains, including the relation of cholic acid to cholesterol, was conclusive. As mentioned below the connexion between ergosterol and vitamin D was appreciated in 1926, and the increase of interest in this aspect of the subject probably led to the next step in advance.

In 1932 Bernal examined ergosterol crystals by means of X-rays and pointed out that the molecular dimensions could not be reconciled with the Wieland-Windaus structures. He found that the molecule was longer and thinner than those theories suggested, and this recalled similar conclusions of Adam & Rosenheim, drawn in 1929 from the study of surface films. In some cases these authors assumed a tilt in the molecules, but they were in general disposed to state that their results could not be explained on the basis of the accepted formulae.

The nettle was grasped by Rosenheim & King who recalled the work of Diels (1927) on the dehydrogenation of certain sterol derivatives, whereby the hydrocarbon chrysene was formed along with other products. They decided to see what would happen if the formation of chrysene (I) was regarded as significant and not, as heretofore, the result of deep-seated decomposition. This led them in May 1932 to advance the formula (II) for desoxycholic acid.



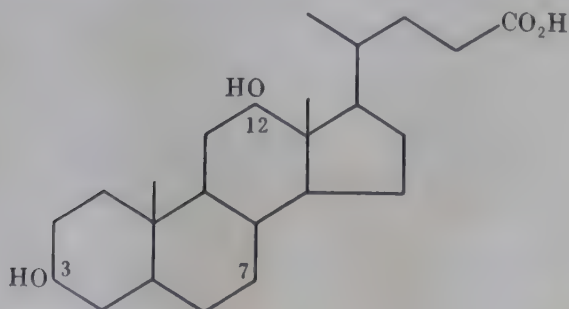
(I)



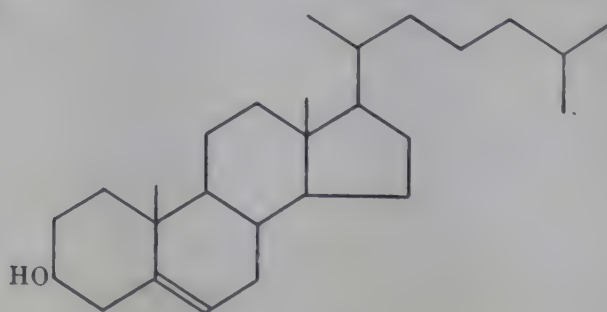
(II)

A structure of this type, applied to ergosterol was found by Bernal to harmonize with his view of the overall dimensions of the molecules.

A few months later in the same year, Wieland & Dane, and also Rosenheim & King, made the final modification to (III) for desoxycholic acid; the corresponding cholesterol structure is (IV) [cholic acid is (III) with a third hydroxyl in position 7].



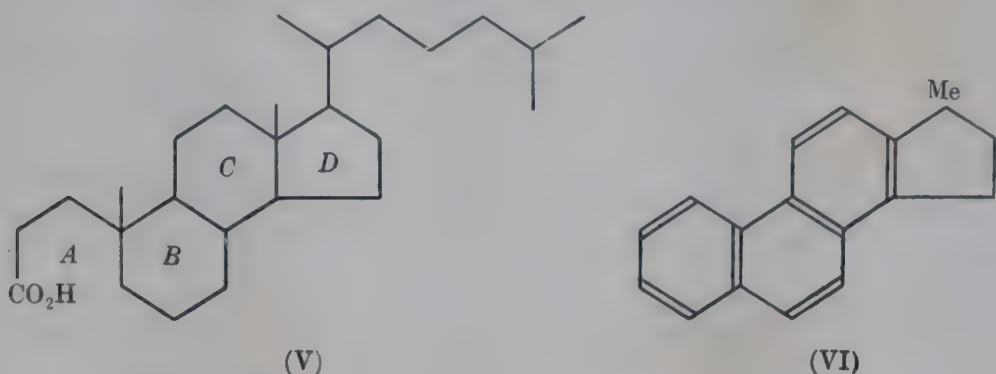
(III)



(IV)

Then followed a busy period in which the assumptions made were tested and found in every case to be valid.

Wieland had shown how to degrade the long side-chain step by step, and Tschesche applied a similar process to an acid (V) obtained by opening ring A.



His results followed the scheme $R \cdot \text{CH}_2 \cdot \text{CH}_2 \cdot \text{CO}_2\text{H}$ (V) $\rightarrow R \cdot \text{CH}_2 \cdot \text{CO}_2\text{H} \rightarrow R \cdot \text{CO}_2\text{H}$ and the properties of this last acid showed it to be of the form $\begin{array}{c} \text{C} \\ \diagup \quad \diagdown \\ \text{C} - \text{C} - \text{CO}_2\text{H} \\ \diagdown \quad \diagup \\ \text{C} \end{array}$ and thus disclosed the position of the side-methyl or angle-methyl group.

In addition to chrysene, Diels obtained a hydrocarbon $\text{C}_{18}\text{H}_{16}$, for example by the dehydrogenation of cholesteryl chloride. This was considered to be (VI), the methyl having wandered into the cyclopentane ring from the angle-position (Cook & Hewitt). This formulation was proved to be correct by synthesis due to Kon, Harper & F. C. J. Ruzicka. An earlier and doubtless successful synthesis by Bergmann & Hillemann was unfortunately not crowned by conclusive comparison with the Diels hydrocarbon. Several other hydrocarbons obtained by dehydrogenation processes have been investigated and synthesized, but by far the most interesting is that known as methylcholanthrene (VIII) because its formation confirms the position of the side-chain and of one of the hydroxyls of cholic acid as well as of the main skeleton.

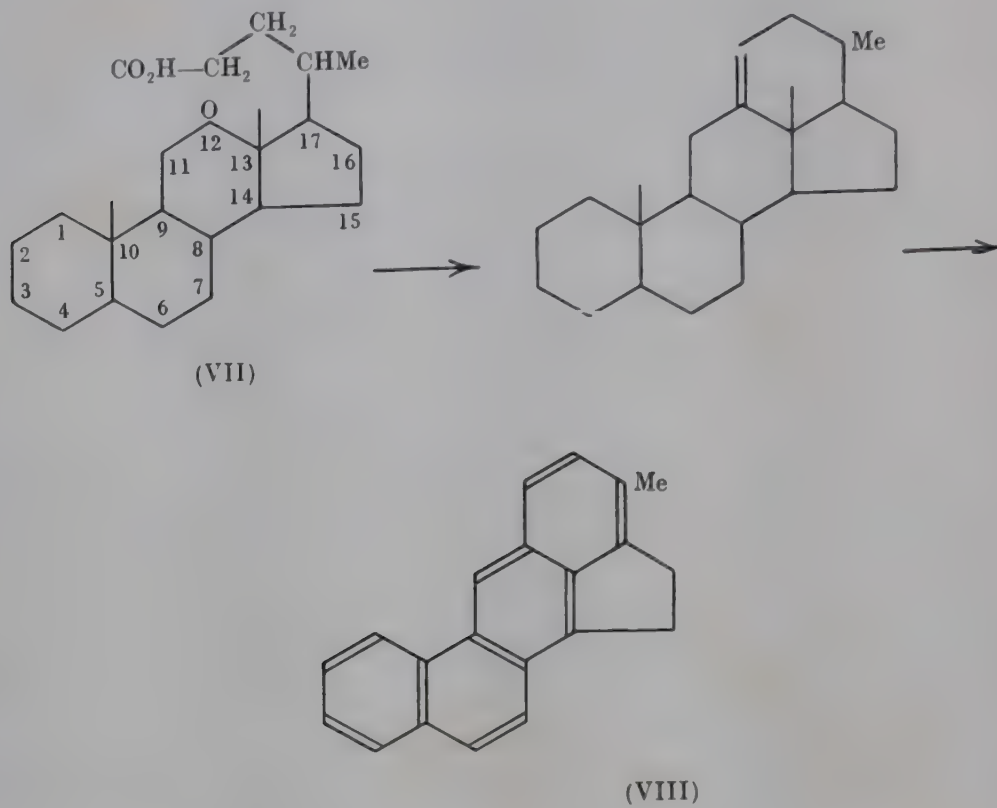
Wieland & Dane (1933) degraded desoxycholic acid to 12-ketocholanic acid (VII) and thence through dehydronorcholene to (VIII).

The last process was independently effected by Cook & Haslewood whose interest was already aroused in 1932 when Kennaway & Cook discussed the possibility of obtaining, from the bile acids, carcinogens analogous to others which they had discovered in the benzanthracene series. Cook & Haslewood demonstrated the constitution of methylcholanthrene by degradation to a tetracyclic compound which they synthesized, and Fieser & Seligman (1935-6) synthesized methylcholanthrene itself in unambiguous fashion.

In passing it may be remarked that there is no positive evidence to connect methylcholanthrene with the etiology of cancer; it has been looked for, and not found, as a constituent of normal and of cancerous tissues.

Thus the long-known facts and newly acquired knowledge fell neatly into place in the light of the new structures and further confirmatory evidence was derived from the total synthesis of oestrogens, namely, equilenin by Bachmann, Cole &

Wilds (1939) and oestrone by Anner & Miescher (1948). The close relation of these substances to cholesterol was always thought to be highly probable but not proven until 1940 when Inhoffen forged the last links in the chain of transformations connecting the two series.



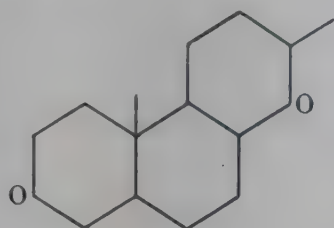
The synthesis of cholesterol itself might well be thought a hopeless quest when it is considered that the formula contains seven asymmetric carbon atoms implying 2^7 stereoisomerides. But one of the doublings is accounted for by resolution into enantiomorphs, and at the worst the problem resolves into six bifurcations in the synthetic route, at each of which some progress must be possible along the right path. We have set ourselves this task and have adopted two methods symbolized as $ABC \rightarrow D$ and $A \rightarrow BCD$, indicating the ring to be added last. Both routes have been developed but we have gone further along the former path. Following it, J. W. Cornforth has provided the first synthetical proof of the correctness of the carbon skeleton in rings *A*, *B* and *C*.

The tricyclic diketone (IX) was obtained by Reich (1945) as a degradation product of desoxycholic acid by way of several intermediates. The Reich diketone can also be obtained from a keto-acetate (X) which was isolated by Köster & Logemann from the debrominated products of the oxidation of cholesterol acetate dibromide.

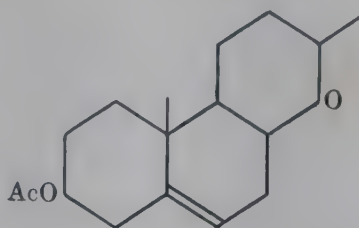
Naphthalene has been transformed into (IX) through nineteen stages and the diketone was obtained in optically active form stereochemically identical with Reich's product.

Although (IX) is obtainable from (X) the reverse operation has not yet been effected. Nevertheless, (X) is much the more readily accessible as a relay, and we are therefore studying the addition of ring *D* to it. The first possible bifurcation has given two substances to H. Holtermann and one of them must possess the correct

orientation. This leaves only one more fork to be negotiated so as to reach the full tetracyclic system in correct stereoisomeric form. We shall be unlucky if the right road is barred here or if it proves impossible to convert (IX) into (X). The long ascent is all but finished at the tetracyclic stage. This constitutes a veritable platform, we can walk round to it, and tackle the few remaining pitches whenever we feel disposed to do so. We are confident that the total synthesis of cholesterol can be realized along these lines.



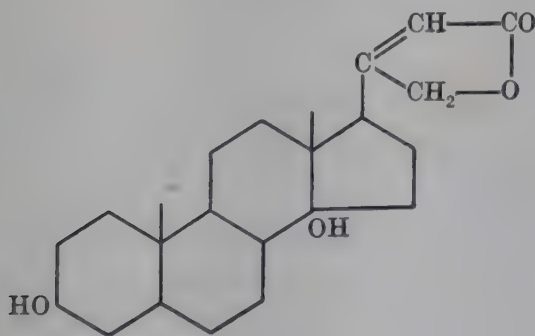
(IX)



(X)

A host of other sterinoids have fallen, or are gradually falling, into line. The constitution of ergosterol, for example, has been demonstrated by the masterly researches of Windaus and of Heilbron. It is impossible to do more than mention the existence of the phytosterols, such as sitosterol and stigmasterol, the very complex saponins, sapogenins, and heart poisons, the sterinoid alkaloids, and so on. All these groups present highly interesting and difficult problems and already have a most extensive literature.

- Many of the plant poisons are glycosides and even the sugars derived from them are unique. One of the simplest, obtained from the purple foxglove, is digitoxin, which yields on hydrolysis, digitoxigenin (XI) and three molecules of digitose (XII) which has a *d*-ribose type of configuration (*d*-desoxy-altrose).



(XI)



(XII)

Another peculiarity of some of these sugars is that they include *O*-methyl groups of true ether function.

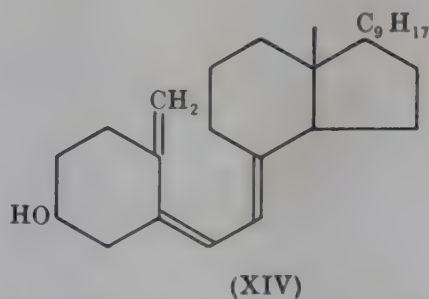
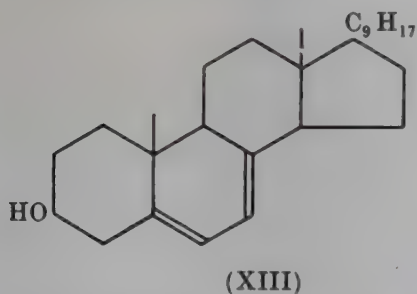
Throughout the chemical work on sterols and their derivatives the problem of stereochemical configuration has been prominent. Much had been done by consideration of transformations, relations and analogies and this was greatly supplemented and extended by X-ray study of crystals. The most exact work in this field has been carried out by Mrs Hodgkin (née Crowfoot) and her collaborators. Their complete study of cholesteryl iodide (Carlisle & Crowfoot 1945) confirmed conclusions drawn from earlier work and clarified all the stereochemical relations of

the molecule. In cholesterol the hydroxyl group and the two angle-methyls lie on the same side of the general plane of the structure. That there is no Walden inversion in the conversion of cholesterol to its iodide may be inferred from the crystal analyses of the iodo-nitrobenzoates of calciferol and lumisterol.

To-day, we have all but a full knowledge of sterol configuration and the results are correlated with rotatory powers as well as with physiological activity.

Vitamin D

Following up the knowledge that rickets could be benefited by sunlight or by addition of certain fish oils to the diet, the active constituent in the oils was concentrated and recognized as a vitamin. Hess, and also Steenbock, found in 1924 that exposure of inactive oils to sunlight or to ultra-violet radiation enhanced their antirachitic properties. Furthermore the pro-vitamin was located in the sterol fraction of the oils. It was found that cholesterol was not this pro-vitamin but that irradiation of ergosterol under suitable conditions produced an antirachitic substance which was termed vitamin D. Intensive study of the transformation products of ergosterol by groups of workers at the National Institute for Medical Research and at Göttingen led to the isolation of the pure vitamin, called calciferol. Then as a result of structural investigations by Heilbron, and by Windaus, and their respective collaborators, the remarkable transformation of ergosterol (XIII) into calciferol (XIV) was established.



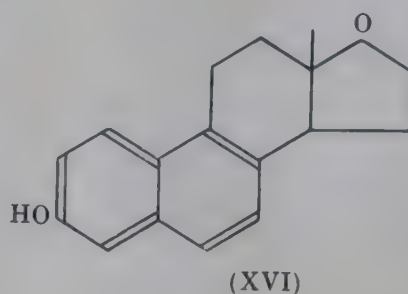
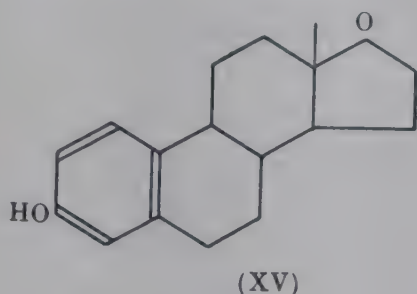
In addition, detailed information about intermediate stages and subsequent changes was garnered and it must be emphasized that this development was possible only as a superstructure on the firm foundations of the constitutional studies of cholesterol and ergosterol. Apart from the side-chain the characteristic feature of the latter is the extra double bond (Δ^7) in ring B and Windaus prepared a cholesterol analogue, namely, 7-dehydrocholesterol and found that it also gave an anti-rachitic substance on irradiation. This was later recognized as the true naturally occurring vitamin, now designated as D_3 . It has been isolated from tunny liver oil by Brockmann and identified, as the 3:5-dinitrobenzoate, with the product from irradiated 7-dehydrocholesterol ((XIII) with side-chain C_8H_{17}). The method of preparation of the pro-vitamin has been greatly improved and it has been found in Nature, for example in the skin of pigs.

Oestrogenic, androgenic, and progestational hormones

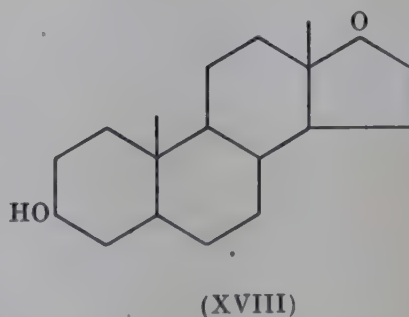
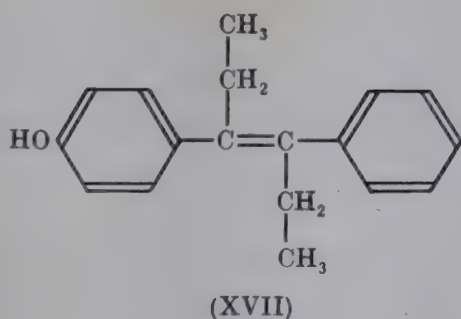
No attempt to cover this field even by way of summary can be made.

It will suffice to notice certain features that connect with the general sterol narrative. The isolation of the first oestrogen, namely, oestrone (XV) by Doisy

(August) and Butenandt (October) occurred in 1929 and it will therefore be appreciated that the constitutional study was largely independent of that of the sterols. The problem was solved by the work of Marrian, of Butenandt, and of Cook from about 1932-4 and the outcome was one of give-and-take with sterol chemistry. The discovery of oestrone was quickly followed by that of other oestrogens; oestriol by Marrian in 1930, equilinen (XVI) by Girard in 1932. Reference has already been made to the synthesis of (XV) and (XVI).



These hormones were at first obtained from the urine of pregnant women and later from that of pregnant mares. In 1934 Zondek found that the urine of stallions was a still richer source, but this surprising result is peculiar to equines. These hormones have also been found in plant material; oestrone in palm kernels and oestriol in female willow flowers. The subterranean clover, so important to Australian agriculture, contains oestrogens which have apparently not yet been identified. The oestrogenic property is, however, not very specific and is exhibited by a considerable range of synthetic compounds of which stilboestrol (XVII), hexoestrol, its dihydro-derivative, and dieneoestrol, a dehydro-derivative, are the best known (Doods and collaborators).

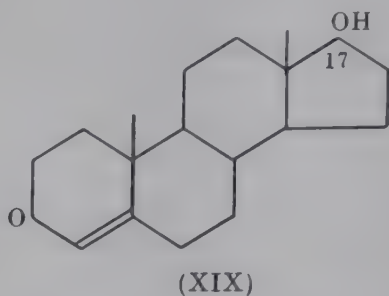


The natural and also the synthetic oestrogens find many applications in therapeutics and both types are manufactured on a considerable scale.

Androsterone (XVIII) was isolated by Butenandt in 1931 from urine of normal males and he almost guessed the correct constitution in 1932. Two years later Ruzicka entered the field with the highly important discovery that he could oxidize away the side-chain of sterol and bile acid derivatives and leave a carbonyl group in its place: $:CH.C_8H_{17} \rightarrow :CO$. The oxidation of cholestanyl acetate gave a weak androgen, similar to, but not identical with, androsterone. Use of a stereo-isomeride, epi-cholestanyl acetate afforded androsterone itself in very small yield.

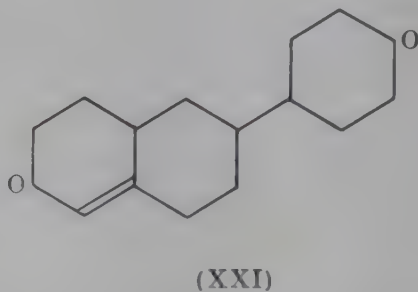
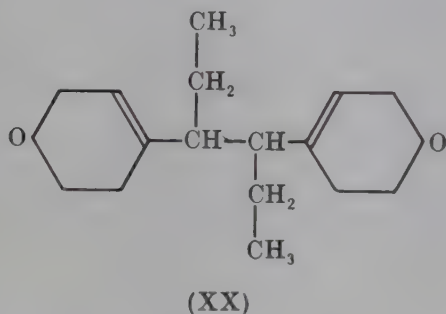
In 1935 Laqueur isolated a very active hormone, namely testosterone (XIX) from testes of steers and this has become the standard androgen in medical practice. It

is manufactured by a development of Ruzicka's method in which the starting point is the dibromide of cholesteryl acetate, which is oxidized by chromic acid. This was reported in 1935 from four different laboratories.



The final stage is the reduction of a carbonyl group in position 17 which is best carried out by a method due to Mamoli; treatment with a sugar-yeast fermenting solution.

Until quite recently no simple synthetic androgenic substance was known, although Birch & Mukherji have described the preparations of XX from hexoestrol by an ingenious method and this substance is under examination for its biological properties.



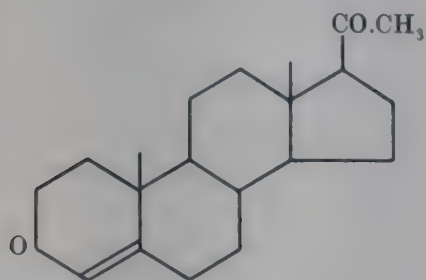
Wilds, Shunk & Hoffman (September 1949) have prepared the diketone (XXI) which is stated to possess approximately 1/200th of the androgenic activity of testosterone in the comb test in day-old chicks. This discovery is of great interest because it gives new encouragement to the search for synthetic analogues of hormones of therapeutic importance.

Progesterone (XXII) is a hormone of the corpus luteum which controls pregnancy. Following indications by W. M. Allen (1932) it was independently isolated in pure form in four laboratories in 1934. The best partial syntheses are from stigmasterol, from various sapogenins, or from a by-product of the oxidation of cholesteryl acetate dibromide.

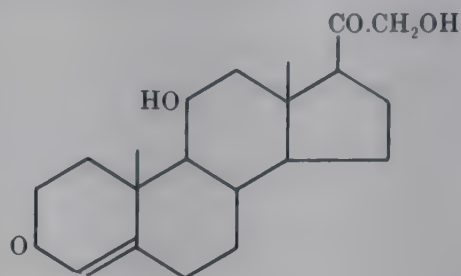
Hormones of the adrenal cortex

Following the indications (1929) that extracts of the gland could prolong the life of adrenalectomized animals, Kendall, Wintersteiner & Pfiffner, and Reichstein have (since 1935) described the isolation of some twenty-eight crystalline substances from the adrenal cortex. This magnificent effort cannot be summarized but the constitutions of all the more important of these substances have been demonstrated by transformations or partial syntheses. Corticosterone (XXIII) was the first active substance to be isolated, first by Reichstein and a little later by Kendall. Like some

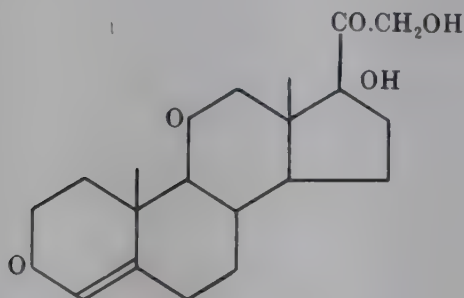
of its congeners it is concerned with regulation of carbohydrate metabolism and electrolyte balance but to discuss these relations in a few words would be impossible. Indeed it is not surprising that some doubt still exists in regard to the precise function of individual hormones in this complex group. The adreno-cortical hormones have been used in the treatment of Addison's disease and surgical or wound shock.



(XXII)



(XXIII)



(XXIV)

At the moment the centre of the stage is occupied by Kendall's substance E (Reichstein's substance F) (XXIV) which is now called cortisone.

Five years ago Selye concluded from his experiments that the adrenal cortex may play an important role in the pathogenesis of rheumatic and rheumatoid conditions in man, and Bassi & Bassi (1946) obtained beneficial effects in chronic rheumatoid arthritis by the administration of adrenal-cortical extracts. Cortisone has now been found by Hench, Kendall, Slocumb & Polley (1949) to alleviate the symptoms of rheumatoid arthritis in a quite dramatic fashion. However, the effect is only temporary and further administration of the hormone is necessary after a certain period. We cannot yet know the length of time during which the improvement can be sustained and undesirable side reactions have appeared in some cases. There are said to be six million rheumatoid arthritics in the United States alone and the provision of cortisone or some effective substitute is therefore a matter of great urgency. But very large quantities will be required and it is doubtful whether the amount needed could conceivably be made by partial synthesis from sterinoid material of natural origin.

At the present time the starting point for successful syntheses is desoxycholic acid which is carried through some thirty-five stages in accordance with brilliant work of Reichstein, and of Sarett, to which many others have contributed by way of modification and improvement of yield. By a remarkable *tour de force* the chemists

of Merck and Co., Rahway, N.J., U.S.A., have made some hundreds of grams of cortisone, but colossal effort was involved. It is claimed that the number of stages would be halved if sarmentogenin could be used as the starting point; this aglycone of a cardiac glycoside from a species of *Strophanthus*, already bears an oxygen atom in position 11. But this material is not available in quantity and, even if it were, the partial synthesis is still too long for the purpose in view. The more promising lines would appear to be (a) total synthesis, (b) synthesis of an effective analogue, or (c) synthesis of a substitute.

The feasibility of finding analogues cannot be assessed in advance. Undoubtedly this idea will be pursued in more than one quarter. Total synthesis is attractive in this case but demands new ideas and new methods; there is a great field for investigation.

The substitute may be found along lines indicated by L. H. Li (First Congress of Biochemistry, Cambridge, 1949). It is first necessary to realize that the hormones already mentioned are released on the arrival of messengers from a higher centre. Thus the anterior pituitary discharges gonadotropic hormones into the blood. These stimulate the gonads to the production of hormones of sterinoid nature. The gonadotropic hormone, studied in this country by Rimington, among others, has protein character. Similarly the anterior pituitary produces adrenocorticotrophic hormone (A.C.T.H.) which stimulates the adrenal cortex and which Hench *et al.* found equivalent to cortisone in the treatment of rheumatoid arthritis. The essence of Li's discovery is that a partial hydrolysate of this protein contains a peptide with a small number of amino-acids (variously stated as 6 to 12) and that this substance is active in the same sense as A.C.T.H. Clearly the synthesis of such a substance might be practicable on a really large scale.

Work on the mode of action of cortisone and A.C.T.H. is of the greatest interest. A recent paper of Selye draws attention to the fact that adrenal cortical hormones are of two main types, mineralo-corticoid and gluco-corticoid and that cortisone and A.C.T.H. are in the latter category. The idea of a balance between these types may explain some apparent anomalies such as the fact that arthritic symptoms do not necessarily appear at a low level of activity of the adrenals.

Finally, mention must be made of some very interesting work by Rittenberg & Bloch on the biogenesis of cholesterol. The sterol is synthesized by surviving rat liver slices *in vitro* from labelled acetate and subsequently degraded in various ways. The results indicate that all the carbon atoms can be derived from acetate. It will soon be possible to attribute to each of the carbon atoms its origin from a methyl or carboxyl of acetic acid.

In spite of, or better, because of, all these brilliant achievements it cannot be supposed that more than a small part of the wonder of the sterols and their derivatives has thus far been brought to light.

Reverting to the tropic hormones of the anterior pituitary, it should be noted that these are by no means the only examples. There seems to be a fairly general and adaptable mechanism whereby the appearance of a protein stimulates the formation

or release of various steroid hormones at different sites in the body. There must be some common factor in these processes and it may not be so far from our reach as we imagine.

There is a widespread belief that steroid hormones are intimately connected in some way with malignancy and for this reason alone research in the field must be vigorously pursued. All the topics we have touched upon reach out into the unknown and all are of profound significance for the progress of biological science and hence for the health and happiness of mankind.

There is a good case for the international organization of a supreme effort which could conveniently be directed in the first instance towards following up the cortisone-A.C.T.H. clues.

Those who control UNESCO have wisely decided to initiate certain International Laboratories, but it is doubtful whether any of the subjects so far advanced, important as they undoubtedly are, should be preferred to a concerted investigation of the chemistry, biochemistry and endocrinology of the sterinoids.

This is not the place or the time to consider the details, which can be adjusted to suit the different circumstances of the participants, but it is already clear that centralization in terms of bricks and mortar would prove difficult.

An organization, with ample funds to be used for the promotion of research in existing centres, and capable of effecting full liaison between them, is the first objective.

The kinetics of the thermal decomposition of normal paraffin hydrocarbons

II. Comparative measurements on the series from propane to *n*-decane

BY F. J. STUBBS AND SIR CYRIL HINSHELWOOD, F.R.S.

(Received 2 September 1949)

The rates of the nitric oxide-inhibited decompositions of hydrocarbons in the series propane to *n*-decane (which according to the results of Part I represent the chain-free molecular reactions) have been measured over a range of pressures. As inferred from the 'apparent chain length' the molecular rearrangement process increases in importance relatively to the chain reaction with the ascent of the series, and, for a given hydrocarbon, with increasing pressure.

For each hydrocarbon, the order of reaction varies between the first and the second. The results are not consistent with a constant order of 1.5 as has been suggested. Nor is the pressure dependence consistent with the uniform transition from second order to first predicted for unimolecular reactions dependent upon a single mode of activation by collision. There appears to be a contribution to the overall reaction from processes which remain of second order up to high pressures.

The decomposition rate for a given hydrocarbon pressure tends to a limit as the series is ascended, for reasons which are discussed.

1. INTRODUCTION

Part I (Stubbs & Hinshelwood 1950) has provided evidence in favour of the hypothesis that the normal reaction by which paraffins decompose is a mixture of radical chain reaction and molecular rearrangement. It is now proposed to study the way in which the rates of these two modes of decomposition vary in the series of normal hydrocarbons from propane to decane.

For reasons already stated it will be assumed that with suitable quantities of nitric oxide all the chains are suppressed. The magnitude in terms of which the velocities will be measured is the initial rate of pressure increase (dp/dt), whether for the normal or for the inhibited reaction. For any given hydrocarbon this is a satisfactory measure of the rate since the following conditions are satisfied. (a) The values of Δp_∞ have been shown to be exactly the same for the normal and inhibited reactions in the examples of *n*-butane, *n*-heptane and *n*-octane. (b) The Δp -time curves have been shown to be accurately superposable over their whole course for 50, 100 and 150 mm. *n*-heptane when the pressures are expressed as fractions of the initial pressure and suitable adjustments are made in the time scales. The question as to how far the values of $(dp/dt)_0$ are comparable for the various members of the paraffin series will be considered in more detail in § 6.

2. THE RELATION BETWEEN RATE AND INITIAL HYDROCARBON PRESSURE FOR THE NORMAL DECOMPOSITION

The initial rate (r_0) for each paraffin was measured at a number of initial pressures (p_0), and the results are given in table 1. The pressure range investigated depended

upon the volatility of the paraffin and varied from 0 to 500 mm. for propane and butane to 0 to 100 mm. for *n*-decane.

TABLE 1. NORMAL DECOMPOSITION AT 530° C

(p_0 in mm. r_0 in mm./min.)

propane:

p_0	100	154	200	245	329	393.5	442.5	525
r_0	0.71	1.25	1.68	2.35	3.27	3.72	4.70	5.35

n-butane:

p_0	57.0	100	154	194	258.5	297.5	354.0	401.0	456	498.5
r_0	1.24	2.49	4.40	5.36	6.80	9.06	10.09	11.24	12.98	15.45

n-pentane:

p_0	29	54	100	146	202	250	333	465
r_0	1.10	1.91	4.20	8.40	12.9	13.2	18.5	30.9

n-hexane:

p_0	25	49	77	100	125	150	180	200	232	269	302	354	402
r_0	1.03	2.57	4.63	8.14	10.7	13.0	15.9	18.5	22.6	26.0	29.9	33.5	39.7

n-heptane:

p_0	23	54	78	99	100.5	103	126	150	163	192.5	239	276
r_0	3.10	4.71	7.90	10.1	10.2	10.3	13.6	16.3	17.0	22.7	30.1	38.7

n-octane:

p_0	20	39	50	60	63	80	100	121	139	160	191
r_0	1.55	4.12	5.46	8.54	9.48	11.5	14.7	21.2	26.8	31.9	41.2

n-nonane:

p_0	17.5	34.5	50	66.5	82.0	100	114	130	152
r_0	2.50	6.55	8.14	12.4	15.5	20.6	23.9	28.8	35.0

n-decane:

p_0	19.5	30.0	37.5	53.0	66.0	77.5	94.0	100	112
r_0	5.15	7.10	10.1	11.7	15.2	20.6	25.1	28.2	30.9

With all the paraffins the variation of r_0 with p_0 could be expressed by an empirical equation of the form

$$r_0 = Ap_0 + Bp_0^2,$$

as shown in figure 1, where the upper line represents the theoretical curve for *n*-nonane and the points indicate the experimental results. The values of the constants *A* and *B* for the series of hydrocarbons are given in table 2.

TABLE 2. CONSTANTS FROM THE EMPIRICAL EQUATION FOR REACTION RATE (NORMAL DECOMPOSITION)

	$A \times 10^2$	$B \times 10^5$		$A \times 10^2$	$B \times 10^5$
propane	0.70	0.67	<i>n</i> -heptane	7.80	22.0
<i>n</i> -butane	2.45	1.00	<i>n</i> -octane	10.1	61.0
<i>n</i> -pentane	3.63	7.33	<i>n</i> -nonane	14.0	61.1
<i>n</i> -hexane	7.03	7.67	<i>n</i> -decane	17.7	98.0

3. THE RELATION BETWEEN RATE AND INITIAL HYDROCARBON PRESSURE FOR THE FULLY INHIBITED DECOMPOSITION

The limiting value (r_∞) to which the initial rate of decomposition falls when fully inhibited was measured for each paraffin at various initial pressures. The results are given in table 3. The variation of r_∞ with p_0 can be expressed by an empirical equation of the same form as that for the normal decomposition, namely, $r_\infty = A'p_0 + B'p_0^2$, and the lower line in figure 1 shows how the experimental points fit the theoretical curve in a typical case. The values of A' and B' for the different paraffins are given in table 4.

TABLE 3. FULLY INHIBITED DECOMPOSITION AT 530° C

(p_0 in mm. r_∞ in mm./min.)

propane:

p_0	100	204	302	402	501
r_∞	0.07	0.19	0.41	0.59	0.84

n-butane:

p_0	96.5	158	198.5	246	300	358	416.5	500
r_∞	0.45	1.03	1.48	2.06	2.47	3.01	4.32	5.36

n-pentane:

p_0	46	100	147	199	246	301	350	404	460	467
r_∞	0.23	0.58	1.14	2.14	2.99	3.83	4.89	7.01	8.20	9.69

n-hexane:

p_0	29	50	72	100	150	203	263	300	357
r_∞	0.60	1.23	1.90	3.51	5.80	7.81	10.6	12.4	16.5

n-heptane:

p_0	25	50	73	100	121	150	180	224
r_∞	0.64	1.60	2.55	3.91	4.60	6.00	8.22	11.3

n-octane:

p_0	25	50	60	80	100	115	141	158	185.5	211
r_∞	0.49	1.34	2.02	2.27	4.07	4.33	6.59	7.00	8.45	12.88

n-nonane:

p_0	15.5	34.0	50.0	70.5	84.0	100	117	147
r_∞	0.52	1.44	2.06	3.09	4.16	5.87	7.04	9.27

n-decane:

p_0	20.0	34.5	50.0	72.5	91.5	100
r_∞	0.82	2.06	3.30	5.15	7.04	9.68

TABLE 4. CONSTANTS FROM THE EMPIRICAL EQUATION FOR REACTION RATE (FULLY INHIBITED DECOMPOSITION)

	$A' \times 10^3$	$B' \times 10^5$		$A' \times 10^3$	$B' \times 10^5$
propane	0.042	0.26	n-heptane	2.70	10.0
n-butane	0.35	1.50	n-octane	2.25	15.5
n-pentane	0.50	2.75	n-nonane	2.60	28.0
n-hexane	2.83	4.70	n-decane	4.20	48.0

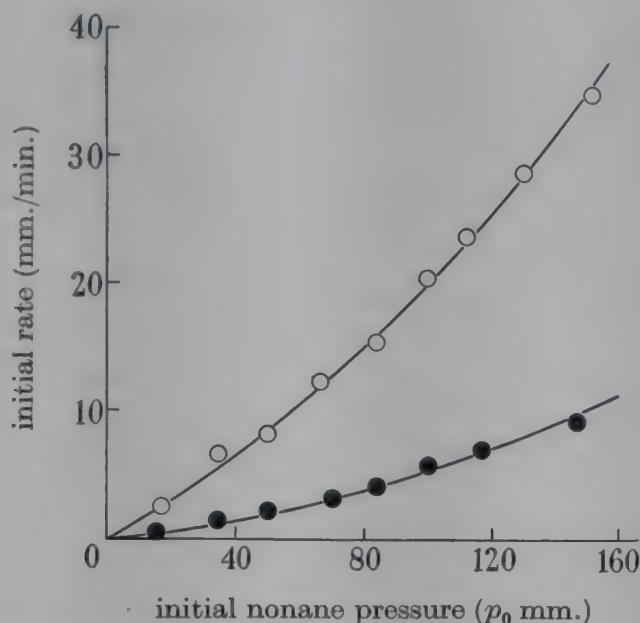


FIGURE 1. Thermal decomposition of *n*-nonane at 530° C. ○, normal decomposition; ●, fully inhibited decomposition.

4. APPARENT CHAIN LENGTH

The ratio r_0/r_∞ is defined as the 'apparent chain length' and serves to measure the proportion of chain reaction. Table 5 gives the variation of apparent chain length with pressure for the different paraffins. For each paraffin it decreases with p_0 , and for a given value of p_0 tends to decrease with ascent of the homologous series. With *n*-pentane at 550° C and with *n*-butane at 560° C the apparent chain length was found to decrease with increasing temperature—as observed by Steacie & Folkins (1940) with *n*-butane.

TABLE 5. APPARENT CHAIN LENGTH AT 530° C

p_0 (mm.)	propane	<i>n</i> -butane	<i>n</i> -pentane	<i>n</i> -hexane	<i>n</i> -heptane	<i>n</i> -octane	<i>n</i> -nonane	<i>n</i> -decane
25	—	—	—	3.0	3.0	4.3	4.4	4.0
50	—	—	6.7	2.5	3.0	4.2	4.1	3.3
75	—	5.4	6.0	2.5	3.0	4.2	4.0	3.3
100	10.0	5.2	5.0	2.4	2.9	4.2	3.7	3.0
150	9.6	4.8	4.8	2.3	2.7	4.2	3.4	—
200	8.9	4.1	4.7	2.3	2.6	4.2	—	—
250	8.2	3.7	4.6	2.3	2.6	—	—	—
300	7.5	3.4	4.3	2.2	—	—	—	—
350	7.0	3.2	4.2	—	—	—	—	—
400	6.6	3.0	4.1	—	—	—	—	—
450	6.3	2.8	4.0	—	—	—	—	—
500	6.1	2.7	3.9	—	—	—	—	—

5. VARIATION OF Δp_∞ IN THE PARAFFIN SERIES

The Δp -time curves for the fully inhibited decomposition of the higher paraffins in the last stages of the reaction became nearly but not quite horizontal. There is a very slow continued rise due to the decomposition of the smaller olefine and paraffin molecules, but the magnitude of this is not great. By a graphical method which,

although somewhat arbitrary in principle, is subject to little uncertainty in practice, the 'end-point' towards which the curve would approach if the slow subsequent decompositions were absent can be estimated. The values of $\Delta p_{\infty}/p_0$ obtained in this way for the various paraffins (100 mm. initial pressure) are recorded in table 6. Experiments with *n*-heptane showed that $\Delta p_{\infty}/p_0$ does not vary with p_0 .

TABLE 6. VARIATION OF $\Delta p_{\infty}/p_0$ FOR THE PARAFFIN SERIES

	$\Delta p_{\infty}/p_0$		$\Delta p_{\infty}/p_0$
propane	0.75	<i>n</i> -heptane	1.75
<i>n</i> -butane	0.93	<i>n</i> -octane	2.00
<i>n</i> -pentane	1.17	<i>n</i> -nonane	2.24
<i>n</i> -hexane	1.44	<i>n</i> -decane	2.52

6. RELATION BETWEEN DECOMPOSITION RATE AND THE NUMBER OF CARBON ATOMS IN THE PARAFFIN MOLECULE

With most of the paraffins secondary decomposition of the product sets in with appreciable speed when enough has accumulated. A rapid enough decomposition, however, of any product, would lead not to a sigmoid curve, but to an initial rate of pressure increase greater by a definite stoichiometric factor than that corresponding to the number of paraffin molecules decomposed in the primary step. This factor might vary from unity in the lower members of the series to a value several times greater for the higher members. Table 6, which shows the variation of $\Delta p_{\infty}/p_0$ for the series, indicates the extreme limit of the possible corrections to $(dp/dt)_0$. In fact, the corrections to be applied are much smaller, and for members of the series up to C_7 are probably negligible. Figure 9 of part I shows that the sigmoid character of the Δp -time curves becomes more pronounced as the paraffin series is ascended. This means that even with the higher members the subsequent reactions of products declare themselves principally after considerable amounts have accumulated and not in the region where the initial rate is measured. With the highest members there might still be need for corrections: decane might give, say, methane and nonane; nonane might decompose immediately to give butane and pentane, the slow subsequent decompositions of the latter imparting the sigmoid form to the curve. The initial rate of pressure increase might thus correspond to a smaller number of paraffin

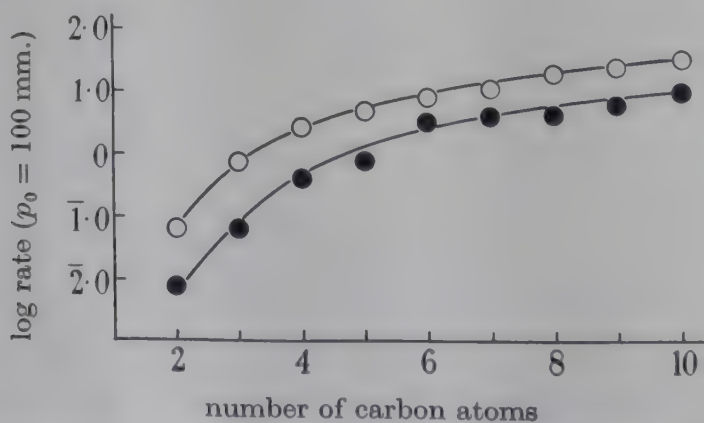


FIGURE 2. Relative reaction rates for members of the normal paraffin series (530° C, $p_0 = 100$ mm.). O, without nitric oxide; ●, with nitric oxide.

molecules than with the lower members. In view of the known reaction rates this would not apply below heptane and is not really likely to apply even to the higher members, since, in fact, the curves are so nearly linear at the start, and such extreme instability would have to be postulated for the higher olefines.

The comparison of initial rates made in figure 2 is therefore based on uncorrected values of $(dp/dt)_0$. If correction is necessary, which is not very probable, it is in the sense that the rates for the few highest members of the series should be reduced, whereby the curve would become still flatter.

Figure 2 shows the plot of $\log(\text{rate})$ for $p_0 = 100$ mm. against the number of carbon atoms for both the normal and fully inhibited decompositions of the series of hydrocarbons. The values for ethane are taken from the results of Partington (1947).

7. DISCUSSION

According to the conclusions reached in part I, the normal decomposition consists of a radical chain reaction and a molecular rearrangement process, the rate of the latter being measured by r_∞ . The chain mechanism is fairly well understood and does not vary a great deal from one paraffin to another (Hobbs & Hinshelwood 1938; Partington, Stubbs & Hinshelwood 1949). For the present discussion the molecular rearrangement process is the more important and will be dealt with in more detail.

The results of § 3 have shown how r_∞ varies with p_0 . It may be assumed that the various values are legitimately comparable with one another, since the Δp -time curves have been shown to be accurately superposable over the whole of their course for 50, 100 and 150 mm. of *n*-heptane when the pressures are expressed as fractions of the initial pressure and suitable adjustments are made in the time scales. Tangents at given points to the curves of $\log r_\infty$ against $\log p_0$ give the order of reaction at the different pressures. The results for the various paraffins are given in table 7. For each paraffin the order increases with increasing pressure, the reaction being almost of the first order at low pressures and tending towards the second at high pressures. Some observers have pointed out that the rate varies approximately as $(p_0)^{1.5}$, and the results given in table 7 show how measurements over a limited pressure range could lead to this result.

TABLE 7. APPARENT ORDERS OF REACTION

p_0 (mm.)	propane	<i>n</i> -butane	<i>n</i> -pentane	<i>n</i> -hexane	<i>n</i> -heptane	<i>n</i> -octane	<i>n</i> -nonane	<i>n</i> -decane
25	—	—	—	—	—	—	1.20	1.05
50	—	1.15	1.25	1.10	1.20	1.25	1.30	1.30
75	—	—	—	—	—	—	1.40	1.45
100	1.15	1.30	1.40	1.15	1.30	1.40	1.65	1.75
150	—	—	—	1.20	1.40	1.50	—	—
200	1.50	1.50	1.55	1.30	1.50	1.65	—	—
300	1.70	1.60	1.60	1.30	—	—	—	—
400	1.80	1.65	1.65	1.40	—	—	—	—
500	1.85	1.70	1.75	—	—	—	—	—

The behaviour predicted for a unimolecular reaction with activation by collision is a transformation from second order at lower to first order at higher pressures. In the simplest case it would follow the expression

$$r = k_1 p_0^2 / (1 + k_2 p_0 / k_3),$$

where k_1 , k_2 and k_3 are constants, whence

$$p_0/r = 1/k_1 p_0 + k_2/k_1 k_3.$$

This equation indicates that the plot of p_0/r against $1/p_0$ should be linear.

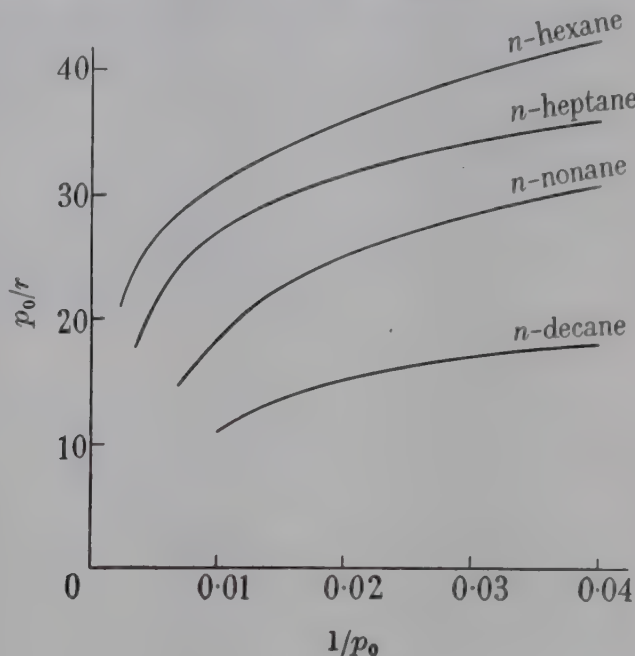


FIGURE 3. Plot of p_0/r against $1/p_0$.

Figure 3 shows that for the paraffin decompositions this is not even approximately fulfilled. (The plots for all the series are similar to the examples given.) It is generally agreed, in accordance with the views of Kassel and others, that k_3 , which measures the probability of transformation of activated molecules, will not be a constant. There are two possible views, not however mutually exclusive, about its variation: (1) k_3 increases continuously with $(E - E_0)$, where E is the energy of the molecule and E_0 a critical minimum for reaction; (2) there is a discrete spectrum of k_3 values, the transformation probability varying according to the particular mode of vibration in which the molecule is excited. Combination of (1) and (2) is quite possible. The intercept of p_0/r on the axis being $k_2/k_1 k_3$, a form of curve concave to the axis of $1/p_0$ indicates that as the pressure falls $1/k_3$ increases, that is, at low values of p_0 there is a greater contribution from small values of k_3 . This is commonly found, since at low pressures molecules escape collisional deactivation for a longer time. The curves in figure 3 are qualitatively of this form, but quantitatively the curvature is so pronounced that the variation which would have to be assumed to explain it would be difficult to account for on hypothesis (1).

The dependence of k_3 on the actual magnitude of the activation energy in a single mode of activation would, in particular, not be sufficient to explain the great predominance of the bimolecular term at large values of p_0 . For this we must suppose that there are other modes of activation in which k_3 is inherently larger, probably because the associated activation energies are small enough. These correspond to a reaction which remains of the second order up to pressures beyond the range of the present experiments. The excitation of these modes of activation would demand the right kind of collision, and provided that this has occurred, the rate of transformation

will be very high. This will explain the occurrence in the reaction velocity expression of a term practically of the classical bimolecular form.

Empirically the relation between rate and initial pressure for the normal reaction is expressed in the form

$$r_0 = Ap_0 + Bp_0^2,$$

and that for the fully inhibited reaction in the form

$$r_\infty = A'p_0 + B'p_0^2.$$

Formally, therefore, it would appear that the suppressible part of the reaction is of the form

$$(r_0 - r_\infty) = (A - A')p_0 + (B - B')p_0^2.$$

The numerical values of the constants are such that for the chain reaction the influence of the term involving $(p_0)^2$ is much less than for the molecular reaction. The quadratic formula therefore may well be no more in this case than an approximation for some other function of p , involving a power between the first and the second. The order of the chain reaction for the decomposition of ethane was dealt with in detail by Hobbs & Hinshelwood (1938). They calculated that the rate would depend upon a power of p_0 which could vary over the range of pressure between one and two. On the assumption that the chain reaction does not vary greatly from one paraffin to the next (as is almost certainly true) then the behaviour seems to be satisfactorily explained.

The values of the apparent chain length show that with increasing pressure the molecular rearrangement process plays a more important role, and that for a given paraffin pressure the same applies generally to the ascent of the homologous series, though here the apparent chain length tends to a limiting value.

As regards the molecular rearrangement process itself there are two matters to consider.

First, the plot of $\log r_\infty$ against the number of carbon atoms in the paraffin shows that the decomposition rate tends to a limiting value. If the correction referred to in § 6 were applied to the higher members of the series the limit would become even more definite.

Secondly, with ascent of the series the ratio B'/A' increases, that is, the contribution to the reaction of the second-order terms increases. The first-order term (A') increases towards a limit, and this is provided for by the standard theory of contributions to the activation energy from an increasing number of internal degrees of freedom. The relative predominance of the second-order term (B') in the higher members of the series is surprising. The simplest hypothesis regarding this seems to be that the energy of activation associated with certain special modes decreases in virtue of structural influences. As the length of the carbon chain increases the central CH_2 groups become further removed from the influence of the terminal groups. Any structural effects which the terminal methyl groups exert on the CH_2 groups will begin to weaken when transmitted over more than two or three carbon atoms. The central part of the molecule is removed by this amount at hexane or octane, and it is beyond this point in the series that the saturation effect begins to be evident. The

modes of activation associated with the bimolecular terms would, according to this consideration, be specially connected with CH_2 groups relatively uninfluenced by CH_3 . The present study therefore gives a clear lead regarding the desirability of obtaining detailed information about the activation energies. Until this is available the discussion is best not carried beyond the present stage.

This work is part of the programme of the Fuel Research Board of the Department of Scientific and Industrial Research and is published with the approval of the Director of Fuel Research.

REFERENCES

- Hobbs, J. E. & Hinshelwood, C. N. 1938 *Proc. Roy. Soc. A*, **167**, 439.
Partington, R. G. 1947 *Disc. Faraday Soc.* 'The Labile Molecule', p. 115.
Partington, R. G., Stubbs, F. J. & Hinshelwood, C. N. 1949 *J. Chem. Soc.* p. 2674.
Steacie, E. W. R. & Folkins, H. O. 1940 *Canad. J. Res. B*, **18**, 1.
Stubbs, F. J. & Hinshelwood, Sir C. 1950 *Proc. Roy. Soc. A*, **200**, 458.

The slow oxidation of gaseous methyl ethyl ketone

BY J. BARDWELL AND SIR CYRIL HINSHELWOOD, F.R.S.

Physical Chemistry Laboratory, University of Oxford

(Received 15 October 1949)

The kinetics of the reaction between oxygen and gaseous methyl ethyl ketone (butanone) have been investigated in the range 250 to 450° C. The anomalous temperature coefficient of the rate offers a good example of the transition from a low-temperature mode of oxidation to a high-temperature mode. Similar behaviour has been observed with hydrocarbons and is ascribable to the diminishing part played at higher temperatures by the peroxides, which at low temperatures play a key role in generating branching chains.

In the low-temperature range the rate of oxidation varies steeply with the initial pressure of butanone and is a very unusual function of the oxygen pressure, passing through a sharp maximum at certain pressures above which the rate falls and then tends to become constant. The maximum tends to disappear as the temperature is raised. Increase of the surface/volume ratio causes a diminution in rate, while addition of inert gases has no appreciable influence.

The various complex relationships are well accounted for in their general form by a modification of the theory previously applied to pentane and hexane. This theory postulates a series of reaction steps of maximum simplicity, permitting a slow branching of chains by fission of peroxides. It leads to a rate expression which according to circumstances may become infinite at certain critical concentrations or remain finite throughout. The critical limits may be related to explosions or cool flames. In the application to butanone oxidation the peculiar form of the dependence of rate on oxygen pressure may be accounted for and a general account given of the observed cool-flame limits.

The agreement of the theory with a number of rather unusual facts is considered to be of greater significance than its lack of quantitative accuracy which is ascribed to the approximate nature of its basis.

INTRODUCTION

The reaction between oxygen and gaseous hydrocarbons may proceed by two mechanisms known respectively as the high-temperature mode and the low-temperature mode. In both of these, chain reactions predominate; the high-temperature

oxidation appears to consist of a cycle of relatively simple radical reactions, while the low-temperature process depends upon the chain-propagating properties of suitable peroxides.

In the low-temperature region the reaction rate increases with time to a maximum. This acceleration is believed to be connected with the increase in concentration of a peroxide capable of slow fission to yield radicals which cause the branching of chains. A real induction period due to the presence of traces of inhibitory substances and possibly other unknown causes can, however, be superimposed on this phenomenon and so still further delay the development of the maximum rate.

A relatively simple theory of the low-temperature oxidation of paraffins was developed by Cullis & Hinshelwood (1947) and applied with reasonable success to the slow combustion of pentane and hexane. It incorporates essential elements from various preceding theories, especially those of Semenov (1935) and Egerton (1928). The theory leads to an expression (p. 33) according to which the reaction rate may vary continuously with the reactant concentrations or may tend to become infinite for a given value of one of them. In the latter case the phenomenon of a branching chain explosion limit becomes evident. The 'cool flames' that are frequently observed in low-temperature oxidation may be considered from this point of view.

The complexities of hydrocarbon oxidation necessitate a number of simplifying assumptions in the theoretical treatment and thereby lead to some degrees of uncertainty in the derived relationships. For this reason it is desirable that the validity of the resulting formulae be tested as widely as possible. Attention has therefore been directed to the oxidation of various substituted hydrocarbons.

Of the compounds investigated so far, one that has provided particularly interesting information is butanone (methyl ethyl ketone). The slow oxidation of this substance shows many points of resemblance to that of the simple paraffin but also presents a number of new and remarkable relations between rate and concentration. Furthermore, cool flames with readily measurable composition limits are observed. The interpretation of these results in terms of the proposed theory is the major purpose of this paper.

EXPERIMENTAL

The apparatus and experimental methods were similar to those used in previous work in this laboratory. The reaction vessel was of silica and measured 6 cm. in diameter and 10 cm. in length. The butanone vapour and oxygen were admitted separately to the hot reaction vessel and the course of combustion was followed manometrically. The pressure-time record generally revealed an induction period at the end of which the rate increased rapidly to a maximum. The maximum rate of pressure change ($\rho_{\text{max.}}$) was taken as a measure of the rate of reaction.

Factors influencing the maximum rate

(1) *Temperature.* The effect of temperature on the maximum rate of oxidation of butanone at 50 mm. pressure is illustrated in figure 1.

The rate becomes appreciable at about 250° C, and at first increases with increasing temperature. The relation then becomes more complex; the rate decreases with in-

creasing temperature over a range and then increases again. Similar behaviour, which is in sharp contrast with that shown by most chemical reactions, has been observed in the oxidation of other compounds and provides strong evidence for the existence of two mechanisms of oxidation (Pease 1938; Mulcahy 1949).

(2) *Oxygen pressure.* The dependence of maximum rate on oxygen pressure for a constant initial butanone pressure of 50 mm. is shown in figure 2. It is seen that at most temperatures the rate first rises as the oxygen pressure is increased, then falls again, eventually becoming constant. The 'hump' in the oxygen-dependence curve becomes very pronounced at temperatures in the neighbourhood of 328° C but disappears at about 400° C.

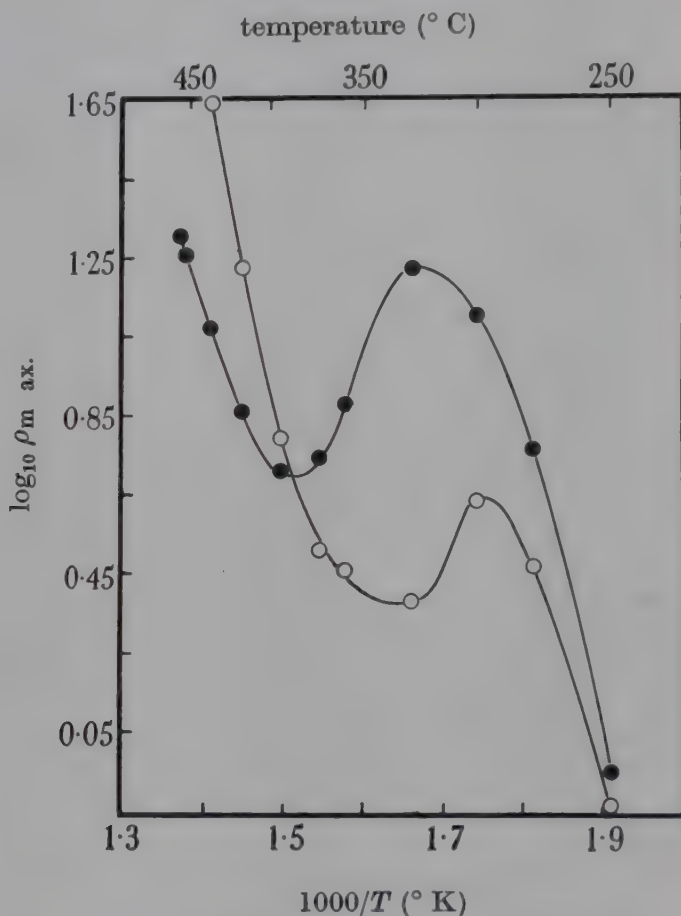


FIGURE 1. The effect of temperature on the maximum rate of oxidation of butanone at: ●, 50 mm., and ○, 400 mm. pressure of oxygen.

The rate-concentration relations at 328° C were studied in detail. Figure 3 shows the variation of maximum rate with oxygen pressure for several initial pressures of butanone. With the higher values of the latter the curves are necessarily incomplete since certain mixtures give rise to cool flames (revealed to experimental observation by brief surges of pressure).

(3) *Butanone pressure.* At all temperatures the effect of butanone pressure on the rate is marked. In figure 4 it is shown for a constant oxygen pressure of 400 mm. at 328° C. The rate rises very steeply, the apparent 'order' of the reaction above about 50 mm. pressure being in the neighbourhood of 6. Above about 100 mm. the combustion is accompanied by cool flames.

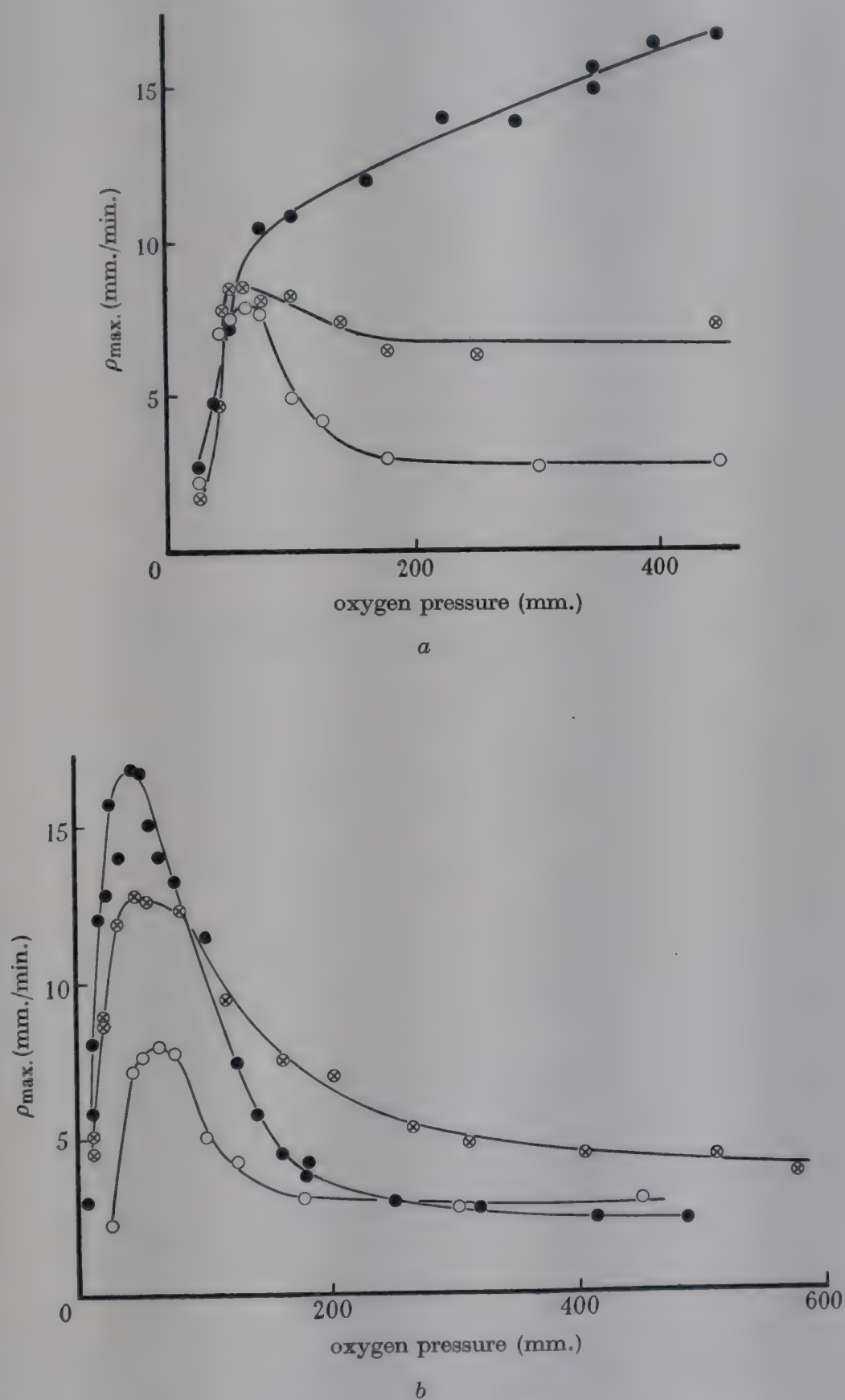


FIGURE 2. The effect of oxygen pressure on the maximum rate of oxidation of butanone at 50 mm. pressure. Temperatures: (a) ●, 417; ⊗, 391; ○, 360° C. (b) ⊗, 300; ●, 328; ○, 360° C.

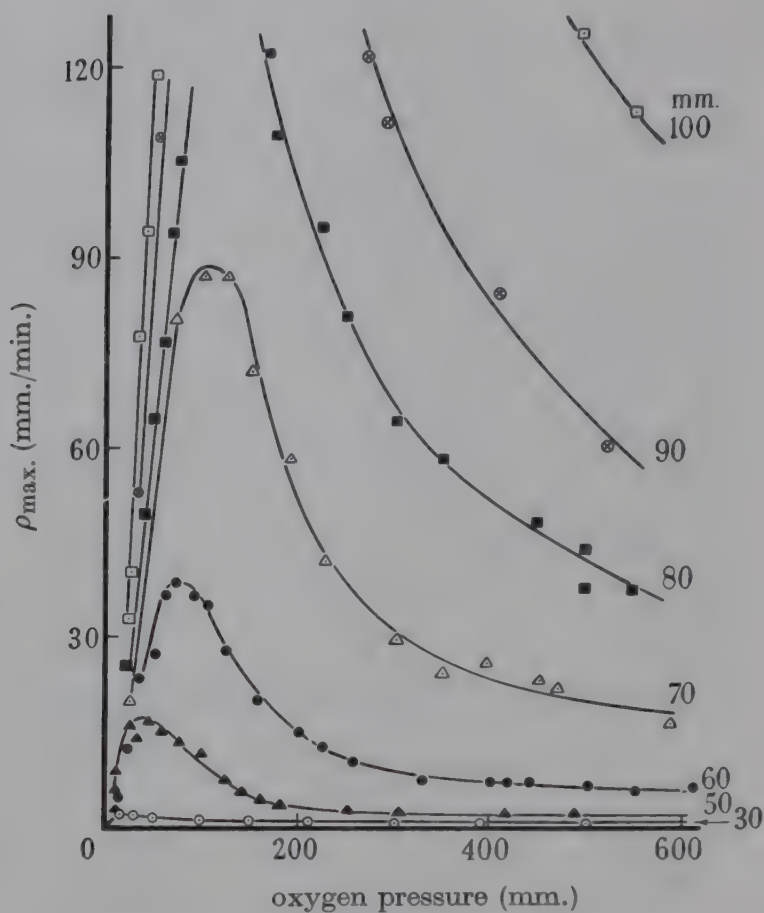


FIGURE 3. The variation of maximum oxidation rate with oxygen pressure for several butanone pressures. Temperature 328°C .

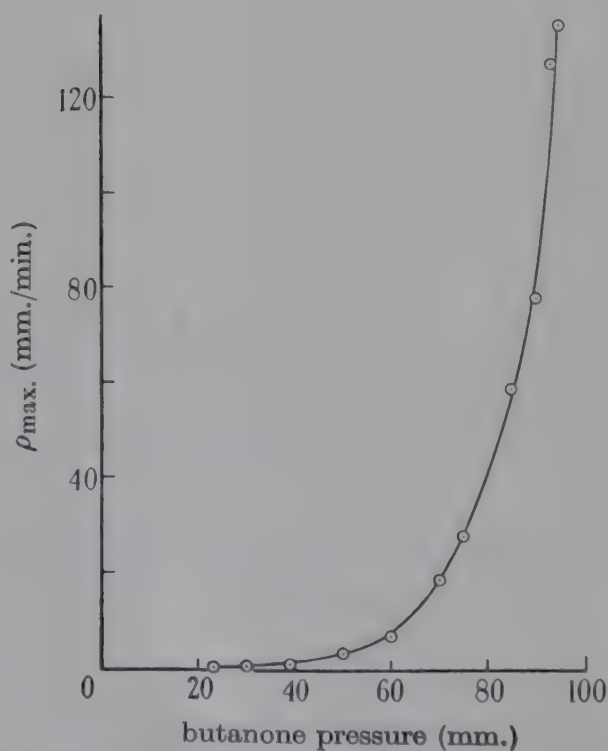


FIGURE 4. The variation of maximum oxidation rate with butanone pressure at 328°C . Oxygen pressure 400 mm.

The butanone pressure at which the oxidation changes over from slow combustion to a cool flame was determined for various other oxygen pressures. The resulting composition chart is shown in figure 5 in which the shaded area represents mixture compositions that eventually give rise to cool flames at 328° C.

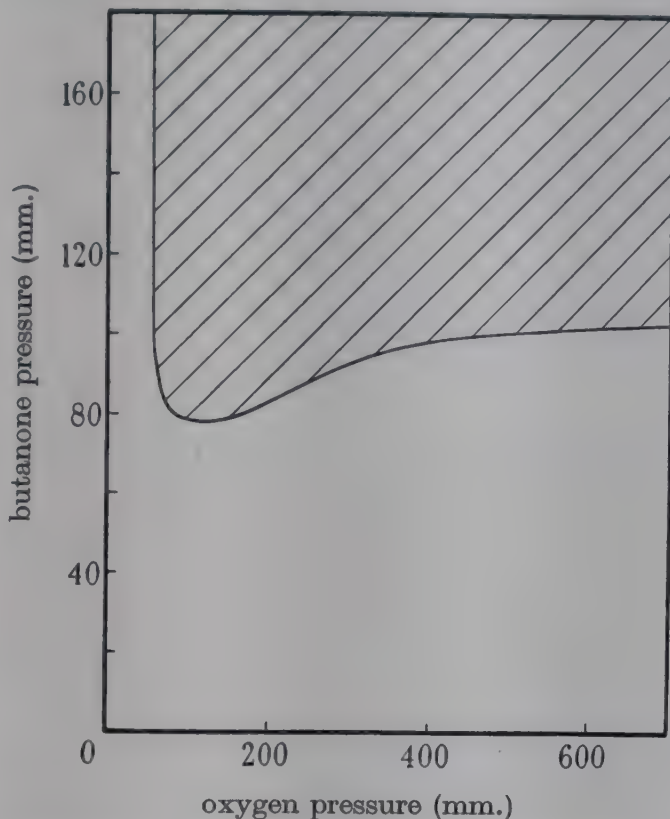


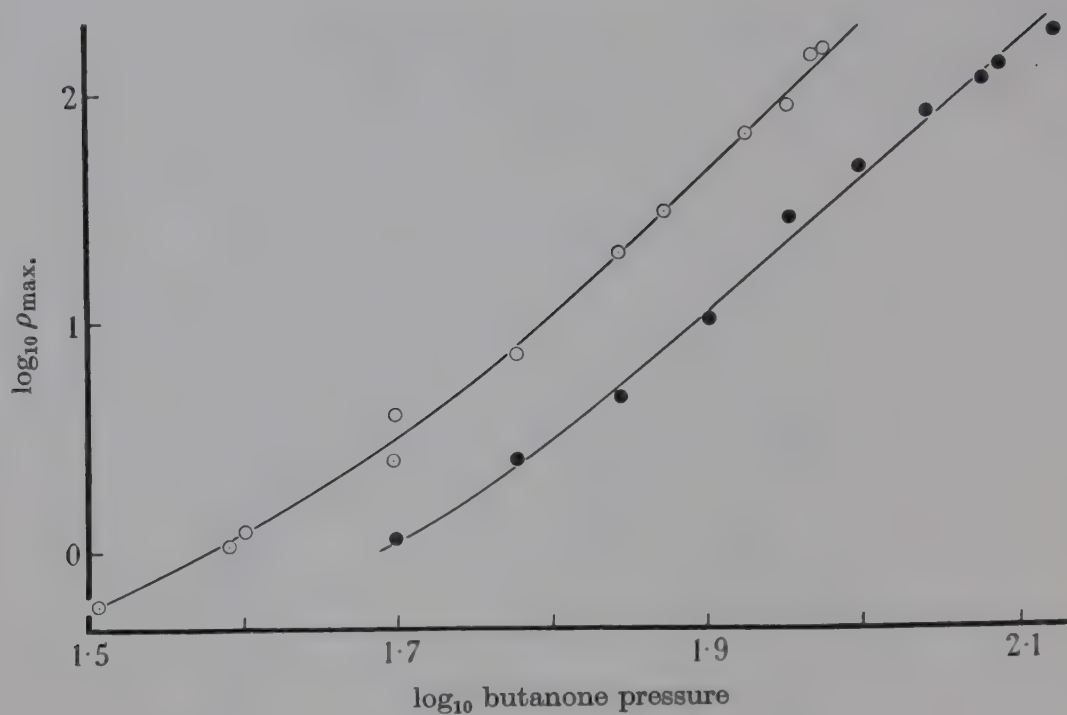
FIGURE 5. Cool-flame limits for butanone-oxygen mixtures at 328° C. The shaded area represents mixtures that lead to cool flames.

(4) *Surface.* A reaction vessel packed with silica tubes was substituted for the normal vessel, thereby increasing the surface/volume ratio by a factor of about 7. After a number of preliminary experiments concordant results were obtained, some of which are shown in figure 6. It is evident that the packing of the vessel reduces the maximum rate of oxidation, but does not seriously alter the form of the dependence on either oxygen or butanone pressure. The net effect of the packing in this instance is roughly equivalent to that of a reduction in the initial pressure of the butanone.

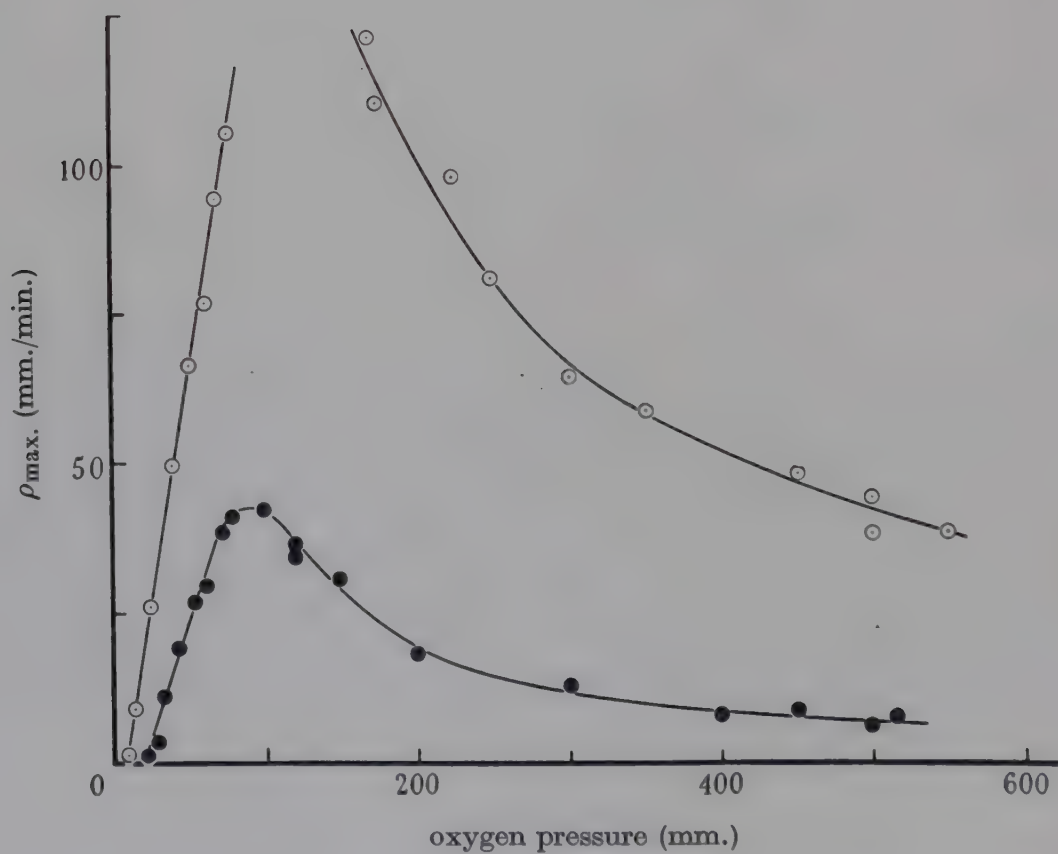
(5) *Inert gases.* Addition of nitrogen or of carbon dioxide to various butanone-oxygen mixtures at 328° C was found to have no significant effect on the maximum rate, either with the packed or with the unpacked vessel.

DISCUSSION

It is evident that in the low-temperature oxidation of butanone the rate depends in a remarkable way on the concentrations of the reactants. These relations are particularly interesting since they are found in their most pronounced form in the region of temperature where most hydrocarbon derivatives react explosively with oxygen. It is therefore of some interest to investigate whether present theories of oxidation are adequate to account for the peculiar relations observed.



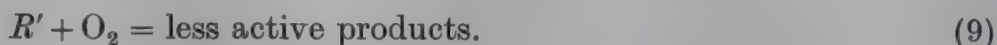
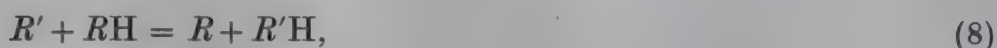
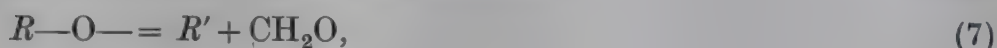
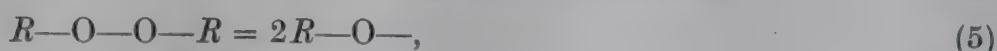
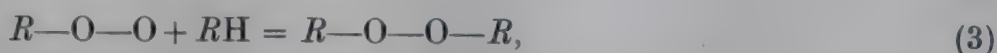
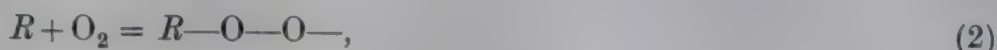
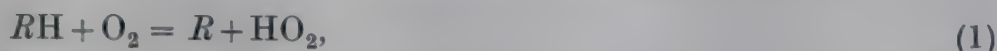
a



b

FIGURE 6. The effect of a sevenfold increase of surface/volume ratio on the maximum oxidation rate of butanone at 328° C. O, unpacked; ●, packed vessel. (a) Oxygen pressure 400 mm., butanone varied. (b) Butanone pressure 80 mm., oxygen varied.

In the theory of low-temperature oxidation proposed by Cullis & Hinshelwood (1947), it is assumed that the essential intermediate is a peroxide, and that the characteristic kinetic features of the reaction are a consequence of the chain-branching that follows fission of the peroxide molecule. The main steps in this scheme are as follows:



Reaction (9) is introduced on the basis of the experimental fact that the rate of oxidation diminishes in a most pronounced way as the length of the hydrocarbon chain becomes shorter. Reaction (7) accounts for the copious and early formation of formaldehyde.

It may be shown that on establishing the stationary state equations for the concentrations of the radicals R , R' and $R-O-O-$ and solving the differential equation for the rate of formation of $R-O-O-R$, the following expression for the rate of reaction is obtained:

$$\rho = \rho_{\max.}(1 - e^{-At}), \quad (I)$$

$$\rho_{\max.} = \frac{2k_1 \left(\frac{k_3}{k_3 + k_4} \right) \left(\frac{k_5}{k_5 + k_6} \right) [RH] [O_2]}{1 - 2 \left(\frac{k_3}{k_3 + k_4} \right) \left(\frac{k_5}{k_5 + k_6} \right) \left(\frac{k_8 [RH]}{k_8 [RH] + k_9 [O_2]} \right)}, \quad (II)$$

where A is proportional to the denominator of the expression (II).

$$\rho_{\max.} = \frac{C_1 [RH] [O_2]}{1 - \frac{C_2 [RH]}{[O_2] + C_3 [RH]}}. \quad (III)$$

C_1 , C_2 and C_3 are constants.

In the low-temperature oxidation of butanone unusual relations exist between maximum rate and reactant concentration (figures 2, 3 and 4). The extent to which the above expression can account for these relations is a crucial test of the validity of the general nature of the reaction scheme proposed.

The peculiar effect of oxygen pressure on the maximum rate at 328° C (figures 2 and 3) is a suitable starting point for the correlation of theory and experiment. The form of the theoretical relation between $\rho_{\max.}$ and oxygen pressure is governed largely by the *relative* values of the constants C_2 and C_3 in equation (III). Rates calculated from equation (III) with $C_1 = 10^{-5}$, $C_2 = 100$ and various values of C_3 are shown in

figure 7. It is seen that if C_3 is greater than C_2 the rate is finite for all oxygen concentrations. If, however, C_3 is less than C_2 the theory predicts branching chain explosions for *all* values of oxygen pressure up to a certain critical limit. Beyond this value the rate decreases, passes through a shallow minimum and then slowly increases.

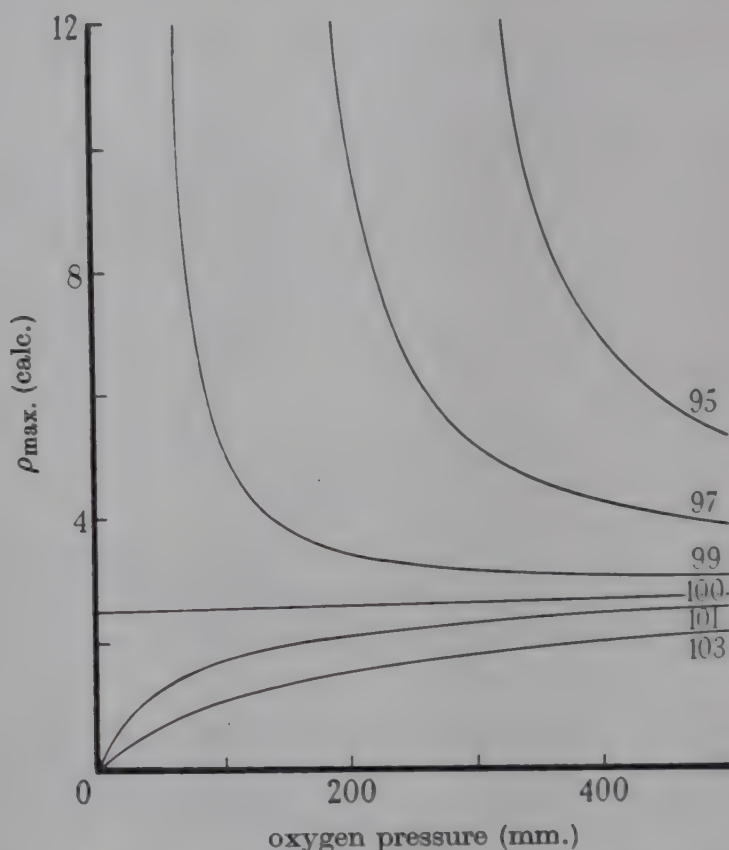
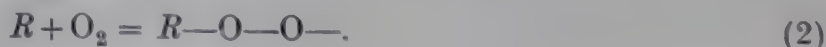


FIGURE 7. Theoretical relation between maximum rate and oxygen pressure according to equation III with $C_1 = 10^{-5}$, $C_2 = 100$ and C_3 as indicated on the curves. Butanone 50 mm.

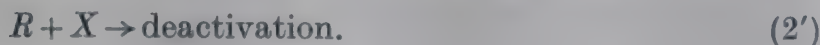
It is clear that in its present form the reaction scheme is incapable of accounting over the entire range for the observed variation of rate with oxygen pressure. In the region of high oxygen pressure, however, the agreement between experiment and theory (for values of C_3 slightly less than C_2) is sufficiently good to suggest that the mechanism is largely correct and that with slight modification it might be adapted to give satisfactory results.

Evidently, in the region of low oxygen pressure, factors not taken into account in the original scheme become important enough to suppress the branching chain explosion predicted by theory. Several possibilities exist. If the active peroxide suffered extensive deactivation at the vessel walls then such deactivation would become of increasing importance at low pressures, and something like the observed falling off of rate might result. If this happened, however, the addition of inert gas would increase the rate by reducing diffusion to the walls. Since no alteration of the rate is observed experimentally when inert gas is added this hypothesis must be rejected.

In the original mechanism only one reaction is permitted for the radical R ; it must react with oxygen according to equation (2):



Other fates of the radical (for example, reaction with product, surface deactivation) have so far been neglected. Such competing reactions must, however, occur to some extent and will now be taken into account. They may for the present be indicated merely by



When equation (2'), which is intended to represent all chain-breaking reactions involving R , is included in the proposed scheme the final expression for $\rho_{\max.}$ becomes

$$\rho_{\max.} = \frac{2k_1 \left(\frac{k_3}{k_3 + k_4} \right) \left(\frac{k_5}{k_5 + k_6} \right) \left(\frac{k_2[\text{O}_2]}{k_2[\text{O}_2] + k'_2[X]} \right) [\text{RH}][\text{O}_2]}{1 - 2 \left(\frac{k_3}{k_3 + k_4} \right) \left(\frac{k_5}{k_5 + k_6} \right) \left(\frac{k_2[\text{O}_2]}{k_2[\text{O}_2] + k'_2[X]} \right) \left(\frac{k_8[\text{RH}]}{k_8[\text{RH}] + k_9[\text{O}_2]} \right)}. \quad (\text{IV})$$

On comparison of this equation with equation (II) it is seen that the inclusion of reaction (2') has resulted in the introduction of a factor

$$\frac{k_2[\text{O}_2]}{k_2[\text{O}_2] + k'_2[X]}$$

in the numerator and in the second term of the denominator of the rate expression.

Figure 8 shows that the presence of this factor can lead to an oxygen-dependence curve similar in form to that found experimentally. The theoretical values of $\rho_{\max.}$ are calculated from equation (IV) with $C_1 = 8.5 \times 10^{-6}$, $C_2 = 100$, $C_3 = 98.5$ and $k'_2[X]/k_2 = 0.32$. The experimental curve is for 50 mm. of butanone at 328°C .

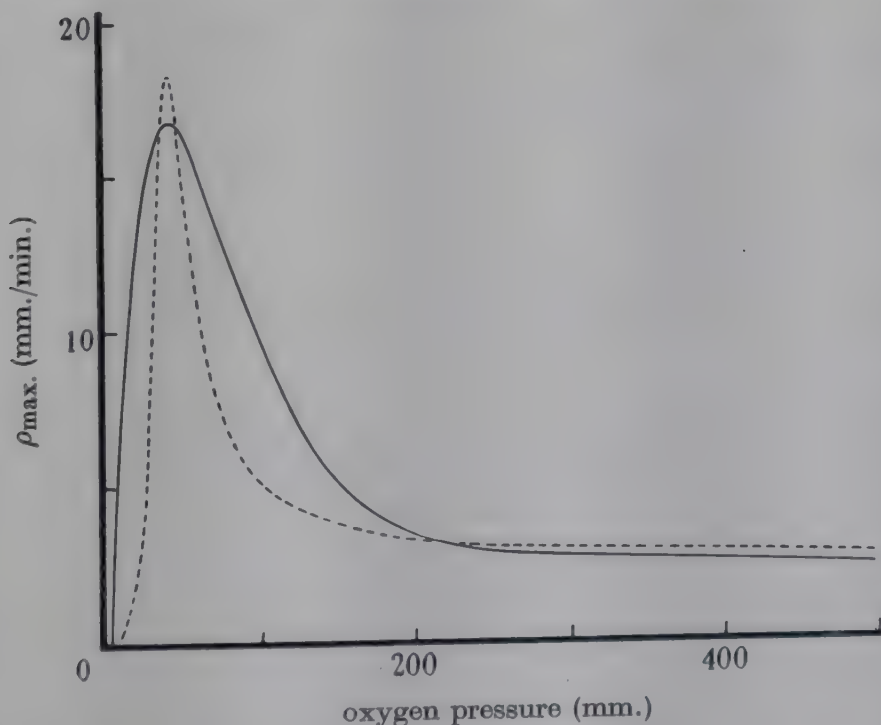


FIGURE 8. Broken curve: theoretical relation between maximum rate and oxygen pressure according to equation (IV) with $C_1 = 8.5 \times 10^{-6}$, $C_2 = 100$, $C_3 = 98.5$ and $k'_2[X]/k_2 = 0.32$. Unbroken curve: experimental for 50 mm. butanone at 328°C .

Although the substitution in equation (IV) of a constant value of the quantity $k'_2[X]/k_2$ does lead to an oxygen-dependence curve similar to that observed, for a

given pressure of butanone, different values must be chosen for this quantity to obtain similar agreement for other butanone pressures. When the *same* value is employed the family of curves shown in figure 9 results. (Only the region of comparatively low oxygen pressure is shown.) Considerable deviation from the experimental curves (figure 3) is found, and it seems likely that better agreement would follow if some variation of the quantity $k'_2[X]/k_2$ with butanone pressure, $[RH]$, were allowed. Possible reasons for such variation will now be considered.

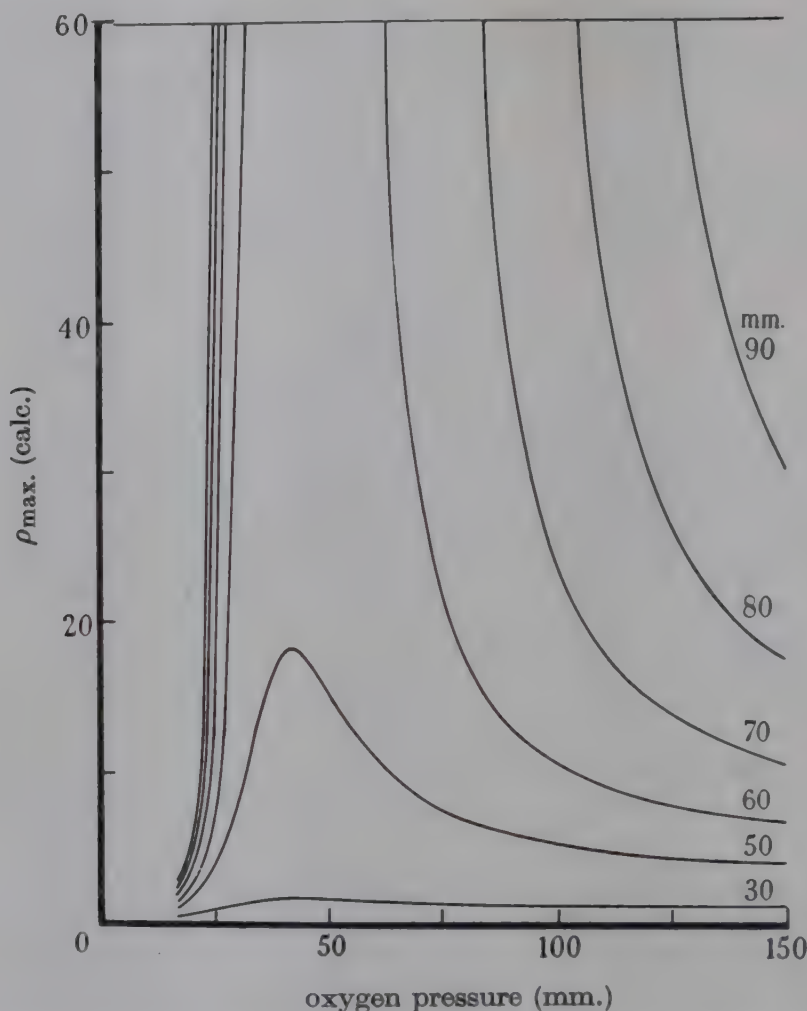


FIGURE 9. Theoretical relation between maximum rate and oxygen pressure according to equation (IV) with $C_1 = 8.5 \times 10^{-6}$, $C_2 = 100$, $C_3 = 98.5$, $k'_2[X]/k_2 = 0.32$, for various pressures of butanone as indicated.

Of the chain-breaking reactions postulated in equation (2') a plausible one is reaction with product; e.g. with formaldehyde which is known in similar cases to be an inhibitor (Cullis & Hinshelwood 1947). The concentration of product at any given stage of the reaction may be expected to vary with the initial pressure of butanone, $[RH]$, and it is in fact found empirically that there is a linear relation between $[RH]$ and the amount of reaction which has occurred up to the moment when the maximum rate is attained. If then as an empirical approximation $k'_2[X]/k_2$ is assumed proportional to $[RH]$, a new family of curves is obtained, illustrated in figure 10 which shows the calculated rates for $C_1 = 8.5 \times 10^{-6}$, $C_2 = 100$, $C_3 = 98.5$ and $k'_2[X]/k_2$ given by $6.4 \times 10^{-3}[RH]$. Although complete agreement with the experimental curves

(figure 3) does not ensue, there is sufficient improvement to suggest that reaction with product may be one of the more important chain-breaking reactions.

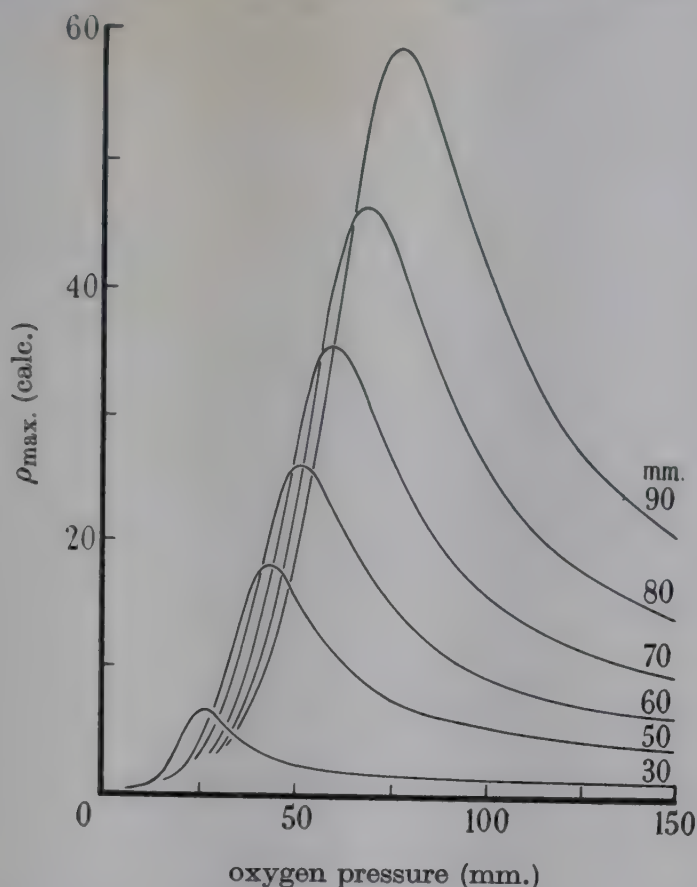


FIGURE 10. Theoretical relation between maximum rate and oxygen pressure according to equation (IV) with $C_1 = 8.5 \times 10^{-6}$, $C_2 = 100$, $C_3 = 98.5$, $k_2[X]/k_2 = 6.4 \times 10^{-3}$ [RH] for various pressures of butanone as indicated.

Dependence of rate on butanone pressure

It has been observed experimentally that with excess oxygen the rate rises very steeply when the butanone pressure is increased, particularly when the composition approaches that leading to a cool flame (figures 4 and 6). Here the apparent 'order' is about 6. The corresponding order of reaction predicted by theory is generally about 2, but rises sharply (and indefinitely) when the butanone pressure approaches that required for a branching chain explosion. In a qualitative way the theory seems to be in harmony with the experimental results, but is not entirely satisfactory from a quantitative standpoint. For example, the experimental 'order' of 6 is maintained over a much larger range of butanone pressure than would be expected on the basis of the theory. Such deviations may, however, result in part from a failure to obtain strictly isothermal conditions in experiments where large rates are found.

Cool-flame limits

When butanone vapour and oxygen are mixed at 328°C certain mixtures give rise to cool flames. The compositions at which these appear are shown in figure 5. It is important to note that the curve shown represents minimum *initial* compositions; the *actual* partial pressures of butanone and oxygen at the moment of inflammation

are somewhat less, as a consequence of the partial consumption of the reactants. Proper consideration of this factor would result in a shift of the curve in figure 5 toward both axes, but it is doubtful if the *form* of the relationship would be much altered.

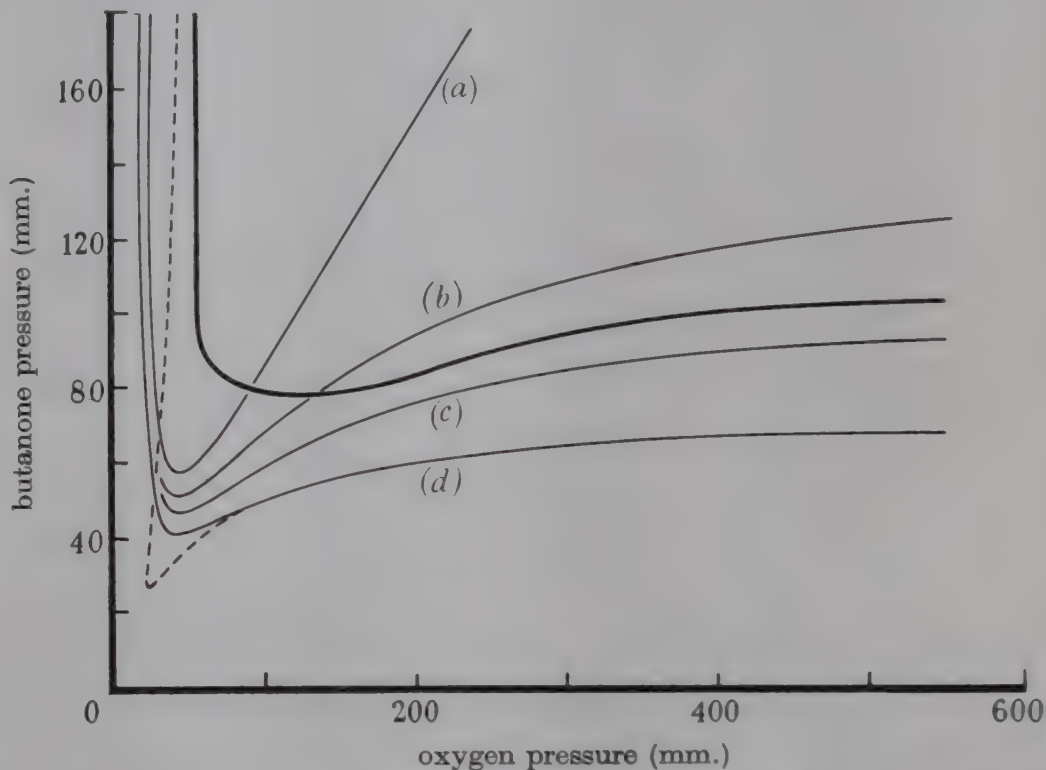


FIGURE 11. Experimental and theoretical cool-flame limits. Heavy curve: experimental at 328°C . The theoretical curves are calculated from equation (IV) with $C_1 = 8.5 \times 10^{-8}$, $C_2 = 100$, $C_3 = 98.5$ and $k'_2[X]/k_2 = 0.32$ and show the compositions required to attain the following rates: curve *a*, ∞ ; *b*, 20 mm./min.; *c*, 10 mm./min.; *d*, 5 mm./min.; broken curve, 5 mm./min. with $k'_2[X]/k_2 = 6.4 \times 10^{-3}$ [RH].

The simplest possible way of bringing the cool flame within the scope of the theory is to suppose that it is a branching chain explosion, and that the observable composition limits correspond to the vanishing of the denominator in equation (IV). In figure 11 curve *a* represents these composition limits calculated with the same constants as were employed in obtaining the curves in figure 9. The departure of the experimental curve from the calculated course suggests that the cool flame can occur in circumstances where the rate is still finite. A similar conclusion might be drawn directly from the results shown in figure 3, and is in agreement with the suggestion of Neumann (1938) that a critical concentration of peroxides is necessary for the appearance of a cool flame.

If then, as the criterion of inflammation, we take, not the complete vanishing of the denominator in the rate expression, but the attainment of various assigned velocities, the series of curves shown in figure 11 *b*, *c* and *d* is obtained. These approximate much more closely in form to the experimental curve and thus support the conclusion that the cool flame results from the attainment of a definite finite rate.

If, instead of the constants used in figure 9 we employ those of figure 10, the composition limits shown by the broken curve in figure 11 are obtained. The same general form of relationship is seen to result.

The effect of surface

In the oxidation of butanone at 328° C the chief effect of increasing the surface/volume ratio is a general reduction in the maximum rate (figure 6). In relating this effect to the reaction mechanism proposed it should be remembered that the reduction in rate is not accompanied by any marked alteration in the form of the rate-oxygen pressure relation. The reduction of the overall rate of reaction is accounted for if we assume an increase in the rate of a step giving rise to products that are inactive in the chain propagation; for example, reaction (4), (6) or (9). Any effective increase in the rate constants k_4 or k_6 would be revealed as a *decrease* in the value of the constant C_2 in equation (III), and it appears from figure 7 that such a change would seriously alter the form of the oxygen dependence. On the other hand, an increase of k_9 would decrease *both* C_2 and C_3 and, although again the rate would be reduced, the form of the oxygen dependence would not be altered much. Consequently, if the effect of surface is to be explained in terms of the scheme suggested, it would appear that reaction (9) is most affected.

CONCLUSION

In the previous sections it has been shown that the theory of slow branching can, in a general way, account for many of the kinetic features of the slow reaction between butanone and oxygen. In view of the approximations inherent in the theory, complete quantitative agreement with experiment can scarcely be expected. Nevertheless, the experimental relations are so unusual in form that even qualitative agreement may be regarded as highly significant.

One of the authors (J.B.) is the holder of a Canadian Postdoctorate Fellowship awarded jointly by the National Research Council of Canada and the Ramsay Memorial Fellowships Trust.

REFERENCES

- Cullis, C. F. & Hinshelwood, C. N. 1947 *Disc. Faraday Soc.* **2**, 117.
Egerton, A. C. 1928 *Nature*, **121**, 10.
Mulcahy, M. F. R. 1949 *Trans. Faraday Soc.* **45**, 537.
Neumann, M. B. 1938 *Acta Physicochimica U.R.S.S.* **9**, 527.
Pease, R. N. 1938 *J. Amer. Chem. Soc.* **60**, 2244.
Semenov, N. 1935 *Chemical kinetics and chain reactions*. Oxford: Clarendon Press.

The contact between a solid and an electrolyte

BY T. B. GRIMLEY

*I.C.I. Research Fellow, Department of Physical and Inorganic Chemistry,
University of Liverpool*

(Communicated by N. F. Mott, F.R.S.—Received 22 June 1949—

Read 3 November 1949—Revised 9 November 1949)

The condition for equilibrium between a binary salt and an electrolyte is that the free-energy change accompanying the transfer of an ion from the interior of the crystal to its hydrated state in solution is zero. In an arbitrary electrolyte this condition can only be satisfied if the two phases are charged. The charge in the electrolyte occurs, at least in part, as a space charge of hydrated ions, the charge density being highest near the interface and falling exponentially to zero in the bulk of the electrolyte. This charged layer, which is well known in colloid science as a factor determining the electrokinetic behaviour of the solid, is usually assumed to be balanced by an opposite charge carried by the crystal in the form of ions adsorbed to the surface. If, however, lattice defects are present in the crystal in thermal equilibrium, the balancing charge may reside actually inside the crystal in the form of a space charge of lattice defects whose structure is similar, in many respects, to the charged layer in the electrolyte. The charge density is highest near the interface and falls exponentially to zero inside the crystal; only at the isoelectric point where the two phases are uncharged are the concentrations of defects uniform throughout the crystal. In any other electrolyte the equations governing the distribution of defects in the crystal are similar to the equations of the Gouy-Chapman theory of the space charge in the electrolyte.

This theory of the double layer is developed for plate-like crystals, and equations are derived which relate the potential drop in each phase and the total charge on the double layer to the physical constants of the system. As the thickness of the crystal is reduced, the space charges at opposite faces begin to overlap in the interior of the crystal so that (except at the isoelectric point) there is no place in the crystal where the defect concentrations are identical with those which exist in thermal equilibrium in an isolated crystal. Because of this overlapping, the thickness of the crystal appears as a parameter in the argument.

The theory is applied to silver bromide in contact with an electrolyte. The defects in the crystal are assumed to be of the Schottky type, so that the charge in the crystal arises through the presence of vacant cation and anion sites in unequal concentrations. On the silver side of the isoelectric point, vacant anion sites are in excess, whilst on the bromide side the reverse is true. c_A^0 , the silver-ion concentration at the isoelectric point, is given by

$$2kT \ln (c_A^0/c_0) = W_B - W_A,$$

where c_0 is the silver (or bromide) ion concentration at the equivalence point, W_A is the work to transfer a silver ion from the interior of the crystal to the electrolyte when the two phases are uncharged and W_B is the corresponding quantity for bromide ions. The double-layer potentials and the total charge are computed as functions of the p_{Ag} for the case in which the electrolyte contains either added silver salts or added bromides only.

The existence of the space charge of vacant sites means that the self-diffusion coefficients of the ions in the crystal are functions of position, and this has an important effect on the rates of exchange of radioactive silver and bromide ions between the electrolyte and the crystal. The rates of exchange are functions of the p_{Ag} , and for silver ions the half-life of the exchange increases as the p_{Ag} increases. For bromide ions the half-life decreases with increasing p_{Ag} . Detailed calculation shows that this phenomenon becomes increasingly important as the dimensions of the crystal are reduced below 10^{-4} cm., and some experimental work with a bearing on this result is discussed.

1. INTRODUCTION

When a solid is in contact with an electrolyte, there generally takes place, at or near the interface, a separation of electric charge; a double layer is set up and the two phases acquire opposite charges. The charge in the electrolyte occurs at least in part

as a space-charge region of free charges in which the charge density is greatest near the interface and falls exponentially to zero in the bulk of the electrolyte (Gouy 1910; Chapman 1913), but although a vast amount of experimental work has already been carried out, the precise nature and location of the balancing charge is not known for any solid/electrolyte system. This state of affairs is undoubtedly due in part to the diversity of systems which are of interest to the physical chemist; in this work we shall be concerned only with the system ionic crystal/electrolyte.

For such systems it has been customary in the literature to consider two ways in which the solid may acquire a charge. First, ions from the electrolyte may be adsorbed on to the surface, and secondly, the surface of the crystal may 'ionize', i.e. the positive and negative ions which are the constituents of the crystal may leave the surface layer and go into solution in unequal numbers (Kruyt & van der Willigen 1928; Verwey & Kruyt 1933; Verwey 1937). We note that in the absence of adsorbable electrolytes, these two mechanisms are indistinguishable. In general, however, both mechanisms are considered to operate, and one is led to consider the existence of a double layer of charge similar in structure to that proposed by Stern (1924). It is no exaggeration to state that, even neglecting the influence of adsorbable electrolytes, no satisfactory theory has yet been put forward which relates the potential drop in the electrolyte to the physical constants of the system. Verwey & Overbeek (1948, p. 46) have attempted to discuss the potential difference between the crystal surface and the interior of the electrolyte (V , say) by assuming that the thermodynamic potential of an ion in the crystal surface is, to a first approximation, independent of V . For a binary salt AB these authors find an expression of the form

$$eV = kT \ln(c_A/c_A^0), \quad (1)$$

where c_A is the concentration of lattice cations in solution and c_A^0 their concentration at the isoelectric point ($V = 0$). Apart from any considerations with regard to the approximations made, the physical significance of (1) is obscure; the only potential difference to which any physical meaning can be attached is a potential difference between two points in the electrolyte (Guggenheim 1929). Whilst it is possible with the aid of certain assumptions to deduce an equation of the form (1) for the potential difference between two points in the electrolyte, one just outside the crystal surface and the other in the interior of the electrolyte, the question will not be dealt with further here. First, because the nature of the assumptions which have to be made is far from clear, and secondly (and this is the most important reason) because there is, in addition to the two mechanisms already mentioned, a third entirely different way in which an ionic crystal in contact with an electrolyte may acquire a charge.

According to the ideas developed mainly by Frenkel (1926), Wagner & Schottky (1930) and Jost (1933), there exist in a crystal in thermodynamic equilibrium at any temperature above the absolute zero a number of vacant lattice sites. Furthermore, a number of ions will be situated in interstitial positions. In the interior of a large crystal (or in any crystal if one applies periodic boundary conditions) these imperfections occur in pairs so that electrical neutrality is preserved. Thus, for example, a vacant cation site must be accompanied by either a vacant anion site or an interstitial cation. These two types of imperfections are usually referred to as Schottky

and Frenkel defects respectively. Now consider what happens when a crystal, containing, for example, Schottky defects, is immersed in an electrolyte.

Unless the free-energy change accompanying the transfer of an ion from the interior of the crystal to the interior of the solution is zero, lattice ions will pass either from the crystal to the electrolyte or in the reverse direction until equilibrium is established. The crystal thus acquires a charge through the presence of vacant cation and anion sites in unequal numbers. If vacant cation sites are in excess the charge is negative, whilst an excess of vacant anion sites corresponds to a positive charge. In either event the charge is not distributed uniformly throughout the crystal; neither is it confined to the surface layer. The equilibrium state is one in which the charge resides inside the crystal in the form of a space-charge region of vacant sites whose structure is in many respects similar to the space-charge region in the electrolyte. The charge density is greatest at the crystal surface and falls exponentially to zero inside the crystal.

Similar conclusions may be reached for crystals which normally contain Frenkel defects. Thus, if the defects are vacant cation sites and interstitial cations, the charge in the crystal arises through the presence, in the space-charge region, of vacant cation sites and interstitial cations in unequal concentrations.

According to these ideas the double layer at the interface between an ionic crystal and an electrolyte may consist entirely of two space-charge regions, one in the crystal and the other in the electrolyte. It is the object of the present paper to examine the structure of such double layers in detail, and in particular to show how the potential difference across the two space-charge regions, and hence the charge on the double layer, can be expressed in terms of the physical constants of the system. In § 2 classical statistics is used to obtain the conditions for equilibrium between an ionic crystal, containing defects, and an electrolyte. In § 3 these results are used to investigate the double layer when both phases are infinite in extent, and the theory is generalized in §§ 4 and 5 to include crystals of finite dimensions. Some general remarks concerning the conditions under which the theory may be expected to apply will be found in § 6, and finally § 7 is devoted to an investigation, with numerical calculations, of the system silver bromide/electrolyte.

2. BASIC STATISTICAL FORMULAE

In this section the conditions for equilibrium between a binary salt AB , containing defects, and an electrolyte will be obtained. It will be assumed that the imperfections in the crystal are vacant cation and anion sites only, but the results will be written in such a form that the generalization to the case where interstitial ions are also present can be performed immediately.

Consider therefore a crystal in which there are N lattice sites available for ions of each type. Denote by N_A and N_B the numbers of occupied cation and anion sites and let $f_A(T)$ and $f_B(T)$ be the corresponding partition functions for the states of vibration of the ions. All interactions between the vacant sites being neglected, the ordinary partition function for the N_A cations and N_B anions in the crystal is

$$K(T) = G(N, N_A, N_B) [f_A(T)]^{N_A} [f_B(T)]^{N_B},$$

where G is the number of distinct configurations of the ions in the crystal, i.e.

$$G = \{N!/N_A!(N - N_A)!\} \{N!/N_B!(N - N_B)!\}.$$

It is convenient to introduce λ_A and λ_B , the 'absolute activities' of the ions, and to construct and use the semi-grand partition function in terms of the variables N , T , λ_A and λ_B (see Fowler & Guggenheim 1939, chap. VI). This function $\Gamma(N, T, \lambda_A, \lambda_B)$ is by definition

$$\Gamma = \sum_{N_A, N_B} \{G(N, N_A, N_B) [\lambda_A f_A]^{N_A} [\lambda_B f_B]^{N_B}\}, \quad (2)$$

and the summations can be performed immediately to give

$$\ln \Gamma = N \ln (1 + \lambda_A f_A) (1 + \lambda_B f_B).$$

The equilibrium values of N_A and N_B are now given in terms of λ_A and λ_B by the equations

$$\left. \begin{aligned} N_A &= \lambda_A N \frac{\partial}{\partial \lambda_A} \{\ln (1 + \lambda_A f_A)\}, \\ N_B &= \lambda_B N \frac{\partial}{\partial \lambda_B} \{\ln (1 + \lambda_B f_B)\}. \end{aligned} \right\} \quad (3)$$

Carrying out the differentiations and writing

$$N - N_A = N_A^h, \quad N - N_B = N_B^h,$$

there results

$$\left. \begin{aligned} N_A &= \lambda_A N_A^h f_A, \\ N_B &= \lambda_B N_B^h f_B. \end{aligned} \right\} \quad (4)$$

The final step in the argument is to express λ_A and λ_B in terms of the concentrations of the ions A and B in the electrolyte. This is readily accomplished. Consider an electrolyte containing a total of M ion pairs of which M_A and M_B are ions of types A and B . The remaining $(2M - M_A - M_B)$ ions are left unspecified. In the simplest case they will be H^+ and OH^- , but this will not be true in general. If there are W water molecules present and if p is the number of water molecules in the hydration shell of an ion, then, employing a quasi-crystalline model and neglecting ionic interactions, the semi-grand partition function $\Gamma'(W, T, \lambda_A, \lambda_B)$ for ions A and B in solution is

$$\Gamma' = \sum_{M_A, M_B} \left\{ \frac{[(W/p) - M_r]!}{M_A! M_B! [(W/p) - 2M]!} [\lambda_A w_A p]^{M_A} [\lambda_B w_B p]^{M_B} \right\}, \quad (5)$$

where $w_A(T)$ and $w_B(T)$ are the partition functions for the states of vibration of the hydrated ions and $M_r = (2M - M_A - M_B)$. Owing to the presence of the factor $[(W/p) - 2M]!$ in the denominator, the summations in (5) cannot be carried out. Γ' may, however, be evaluated by searching for the maximum term. This occurs at the values M_A, M_B given by

$$\left. \begin{aligned} M_A &= \lambda_A (W - 2pM) w_A, \\ M_B &= \lambda_B (W - 2pM) w_B, \end{aligned} \right\} \quad (6)$$

and these are the equilibrium values of M_A and M_B . Introducing the concentration in theoretical units by writing

$$M_A/W = c_A, \quad M_B/W = c_B, \quad M/W = c, \quad (7)$$

equations (6) may be written

$$\begin{aligned} c_A &= \lambda_A(1 - 2pc)w_A, \\ c_B &= \lambda_B(1 - 2pc)w_B. \end{aligned} \quad (8)$$

Eliminating the λ 's between (4) and (8) we find

$$w_A/f_A = \{N_A^h/N_A\}\{c_A/(1 - 2pc)\}, \quad (9)$$

$$w_B/f_B = \{N_B^h/N_B\}\{c_B/(1 - 2pc)\}, \quad (10)$$

which are the desired equilibrium conditions.

In case interstitial ions are also present it may be convenient to express the equilibrium in terms of these rather than in terms of vacant sites. We denote the number of interstitial positions by αN , where α is a numerical factor depending only on the crystal structure. Then if N_A^i and N_B^i are the numbers of interstitial ions and $f_A^i(T)$ and $f_B^i(T)$ the corresponding partition functions, we may infer from (6) that the absolute activities λ_A and λ_B satisfy the equations

$$\begin{aligned} N_A^i &= \lambda_A(\alpha N - N^i)f_A^i, \\ N_B^i &= \lambda_B(\alpha N - N^i)f_B^i, \end{aligned} \quad (11)$$

where $N^i = N_A^i + N_B^i$. Equations (11) can of course be verified by introducing the appropriate extra factors into the partition function (2). Elimination of the λ 's between (8) and (11) yields the two equations

$$w_A/f_A^i = \{(\alpha N - N^i)/N_A^i\}\{c_A/(1 - 2pc)\}, \quad (12)$$

$$w_B/f_B^i = \{(\alpha N - N^i)/N_B^i\}\{c_B/(1 - 2pc)\}, \quad (13)$$

which are the equilibrium conditions.

3. INFINITE CRYSTALS

3.1. *The equilibrium conditions.* In this section explicit forms of the partition functions will be introduced into the equilibrium conditions (9) and (10). We shall assume that the defects in the crystal are of the Schottky type and that we have to deal with a large crystal, so that, by going far enough away from the interface, the conditions can be made to approximate as closely as we please to those which would be obtained by applying periodic boundary conditions, i.e. far from the interface the charge density is zero and we may set $N_A^h = N_B^h$.

If the partition function for an ion in the crystal be referred to its lowest quantum state we may set

$$f_A(T) = f'_A(T), \quad (14)$$

where, if the separation of the energy levels is not too small compared with kT , $f'_A(T)$ is of order of magnitude unity. Introducing a quantity Q_A which we define as the work to transfer a cation from the interior of the crystal, across the double layer, to its hydrated state in solution, the partition function for a cation in solution is

$$w_A(T) = w'_A(T) \exp\{-Q_A/kT\}. \quad (15)$$

Here $w'_A(T)$ is the partition function referred to the lowest energy state of the hydrated ion; if the temperature is not too high $w'_A(T)$ is of order of magnitude unity.

Following Fowler (1932) we shall assume that $f'_A(T)/w'_A(T) = 1$, so that inserting (14) and (15) into (9) we find

$$Q_A = -kT \ln \{c_A/(1-2pc)\} - kT \ln \{N_A^h/N_A\}. \quad (16)$$

In a similar way, if Q_B is the work to transfer an anion from the crystal to the electrolyte, (10) yields

$$Q_B = -kT \ln \{c_B/(1-2pc)\} - kT \ln \{N_B^h/N_B\}. \quad (17)$$

In these equations $N_A^h = N_B^h$, so that subtraction of (16) from (17) gives

$$Q_B - Q_A = kT \ln (c_A/c_B). \quad (18)$$

Using the solubility product relation

$$c_A c_B = c_0^2, \quad (19)$$

where c_0 is the cation (or anion) concentration at the equivalence point ($c_A = c_B$), (18) may be put into the form

$$Q_B - Q_A = 2kT \ln (c_A/c_0). \quad (20)$$

We note that Q_A and Q_B in general contain a contribution due to the existence of a double layer of free charges on either side of the interface. However, for one particular cation concentration, which we denote by c_A^0 , the double layer of free charges vanishes. This is the *isoelectric point* of the crystal. The values of Q_A and Q_B under these conditions are important physical constants of the system and we shall denote them by W_A and W_B respectively, i.e. W_A is the work to transfer a cation from the crystal to the electrolyte at the isoelectric point where the two phases are uncharged; W_B is the corresponding quantity for anions. Using (20) we therefore characterize the isoelectric point by the equation

$$2kT \ln (c_A^0/c_0) = W_B - W_A. \quad (21)$$

The quantity Λ defined by

$$2 |e| \Lambda = (Q_B - Q_A) - (W_B - W_A) \quad (22)$$

plays a very important role in the theory of the double layer developed in this paper. It is the contribution which the double layer of free charges makes to the energy difference ($Q_B - Q_A$), and it may be expressed in terms of the potential drop across the two space-charge regions. Thus, if $V_c(0)$ is the potential drop in the crystal, i.e. the potential difference between a point just inside the crystal at the interface and a point in the interior of the crystal, whilst $V_s(0)$ is defined in a similar manner in the solution, then, for monovalent ions A and B , Λ is given by

$$\Lambda = V_s(0) - V_c(0). \quad (23)$$

Furthermore, from (20), (21) and (22) we have

$$e\Lambda = kT \ln (c_A/c_A^0), \quad (24)$$

so that when c_A is given, Λ is completely determined in terms of the physical constants of the system. Thus if a further relation connecting $V_c(0)$ and $V_s(0)$ can be found,

these two quantities will be determined uniquely. This second relation is of course obtained from the condition that the system as a whole is electrically neutral. We require therefore expressions for the total charge in the crystal and the total charge in the electrolyte.

3.2. *The charge in the crystal.* In this section we shall investigate the distribution of vacant sites in the space-charge region in the crystal, and in particular we shall relate the total charge on this layer to the potential drop across it.

With respect to a zero of potential in the interior of the crystal, let $V_c(x)$ be the electrical potential at distance x from the (plane) interface. Then $f_A(T; x)$, the vibrational partition function for a cation at the point x , is given by

$$f_A(T; x) = f_A(T; \infty) \exp \{-eV_c(x)/kT\}. \quad (25)$$

If $N_A(x)$ is the number of occupied cation sites and $N_A^h(x)$ the number of such sites which are vacant, then (cf. equation (4))

$$N_A(x) = \lambda_A N_A^h(x) f_A(T; \infty) \exp \{-eV_c(x)/kT\}. \quad (26)$$

Far from the interface $V_c(x) \rightarrow 0$ as $x \rightarrow \infty$, so that (26) gives

$$N_A(\infty) = \lambda_A N_A^h(\infty) f_A(T; \infty). \quad (27)$$

Eliminating λ_A between (26) and (27) we find

$$N_A^h(x) = [N n_0 \exp \{eV_c(x)/kT\}] / [N - n_0 + n_0 \exp \{eV_c(x)/kT\}], \quad (28a)$$

where n_0 is the concentration of vacant sites in the interior of the crystal, i.e.

$$n_0 = N_A^h(\infty) = N_B^h(\infty).$$

The distribution function for vacant anion sites may be written down by analogy with (28a). Thus

$$N_B^h(x) = [N n_0 \exp \{-eV_c(x)/kT\}] / [N - n_0 + n_0 \exp \{-eV_c(x)/kT\}]. \quad (28b)$$

If the potential drop in the crystal is small so that

$$n_0 \exp \{e | V_c(x) | / kT\} \ll N, \quad (29)$$

equations (28) reduce to

$$\left. \begin{aligned} N_A^h(x) &= n_0 \exp \{eV_c(x)/kT\}, \\ N_B^h(x) &= n_0 \exp \{-eV_c(x)/kT\}, \end{aligned} \right\} \quad (30)$$

but we note that the distributions of vacant sites do not in general follow simple Boltzmann laws. Of course equations (28) are only approximate because we have neglected interactions between vacant sites.

The potential $V_c(x)$ is obtained by solving Poisson's equation

$$d^2V_c/dx^2 = -(4\pi e/\kappa) (N_B^h - N_A^h), \quad (31)$$

κ being the dielectric constant of the crystal. Using the distribution functions (28) and making the transformations

$$\phi_c = eV_c/kT, \quad z_c = \rho_c x, \quad \rho_c = \sqrt{(8\pi e^2 n_0 / \kappa kT)}, \quad (32)$$

(31) integrates under the condition $d\phi_c/dz_c \rightarrow 0$ as $\phi_c \rightarrow 0$ to give

$$n_0(d\phi_c/dz_c)^2 = N \ln \{(N - n_0 + n_c e^{\phi_c})(N - n_0 + n_0 e^{-\phi_c})/N^2\}. \quad (33)$$

The second integration cannot be performed explicitly, but the expression for σ_c , the total charge in the crystal, follows at once from (33). Writing

$$\phi_c(x) = 2 \ln \xi(x), \quad (34)$$

and setting

$$F\{\xi(x)\} = [\{N - n_0 + n_0 \xi^2(x)\} \{N - n_0 + n_0 \xi^{-2}(x)\} / N^2], \quad (35)$$

we find

$$\sigma_c = \sqrt{[(\kappa k T N / 2\pi e^2) \ln F\{\xi(0)\}]} \quad \text{electronic charges.} \quad (36)$$

If the charge in the crystal is small so that (29) is satisfied, we may employ the Boltzmann distributions (30) directly in (31). The equations are now very similar to the well-known equations of Gouy (1910) and Chapman (1913). We shall not reproduce the analysis here (for a recent investigation of these equations see Verwey & Overbeek 1948, chap. II). The final expression for the total charge is

$$\sigma_c = \sqrt{(2n_0 \kappa k T / \pi e^2) \sinh \{\phi_c(0)/2\}}. \quad (37)$$

3.3. The charge in the electrolyte. An expression for this charge may be obtained once satisfactory distribution functions for the ions in the electrolyte have been set up. We consider only 1-1 electrolytes.

With respect to a zero of potential in the interior of the electrolyte, let $V_s(x)$ be the electrical potential at distance x from the interface. Then $w_+^i(T; x)$, the vibrational partition function for the i th type of cation at the point x , is given by

$$w_+^i(T; x) = w_+^i(T; \infty) \exp \{-eV_s(x)/kT\}. \quad (38)$$

If $c_+^i(x)$ is the concentration of the i th type of cation and $\nu(x)$ the total concentration of ions, then by equation (8)

$$c_+^i(x) = \lambda_+^i \{1 - p\nu(x)\} w_+^i(T; \infty) \exp \{-eV_s(x)/kT\}, \quad (39)$$

and far from the interface we shall have

$$c_+^i(0) = \lambda_+^i \{1 - 2pc\} w_+^i(T; \infty), \quad (40)$$

where c is the total concentration of ion pairs in solution. Eliminating λ_+^i between (39) and (40)

$$c_+^i(x) = c_+^i(0) \left\{ \frac{1 - p\nu(x)}{1 - 2pc} \right\} \exp \{-eV_s(x)/kT\}. \quad (41)$$

$c_+(x)$, the total concentration of cations at the point x , is obtained by summing (41) over all i . The distribution function for anions may be found in just the same way.

The results are

$$\left. \begin{aligned} c_+(x) &= c \left\{ \frac{1 - p\nu(x)}{1 - 2pc} \right\} \exp \{-eV_s(x)/kT\}, \\ c_-(x) &= c \left\{ \frac{1 - p\nu(x)}{1 - 2pc} \right\} \exp \{eV_s(x)/kT\}, \end{aligned} \right\} \quad (42)$$

and from its definition

$$\nu(x) = c_+(x) + c_-(x).$$

If the magnitude of the potential drop in the electrolyte is small so that

$$c \exp \{e | V_s(x) | / kT\} \ll 1, \quad (43)$$

equations (42) reduce to

$$\left. \begin{aligned} c_+(x) &= c \exp \{-eV_s(x)/kT\}, \\ c_-(x) &= c \exp \{eV_s(x)/kT\}, \end{aligned} \right\} \quad (44)$$

but we note that, in general, the distributions of ions in the space-charge region in the electrolyte do not follow simple Boltzmann laws. Ionic interactions are not of course included in (42), and the deviations from the Boltzmann laws are associated essentially with the finite sizes of the ions.

The electrical potential $V_s(x)$ satisfies Poisson's equation

$$d^2V_s/dx^2 = -(4\pi eW/D)(c_+ - c_-), \quad (45)$$

where D is the dielectric constant of the electrolyte. Using equations (42) and making the transformations

$$\phi_s = eV_s/kT, \quad z_s = \rho_s x, \quad \rho_s = \sqrt{(8\pi e^2 cW/DkT)}, \quad (46)$$

(45) integrates under the condition $\phi_s \rightarrow 0$ as $d\phi_s/dz_s \rightarrow 0$ to give

$$pc(d\phi_s/dz_s)^2 = \ln \{1 + 4pc \sinh^2(\phi_s/2)\}. \quad (47)$$

The second integration cannot be performed explicitly, but the expression for σ_s , the charge in the electrolyte, follows at once from (47):

$$\sigma_s = \sqrt{[(DWkT/2\pi e^2 p) \ln \{1 + 4pc \sinh^2[\phi_s(0)/2]\}]} \quad (48)$$

When condition (43) is fulfilled, (48) reduces to

$$\sigma_s = \sqrt{(2cWDkT/\pi e^2)} \sinh \{\phi_s(0)/2\}, \quad (49)$$

and the equations governing the distribution of charge in the electrolyte are of course the equations of the well-known Gouy-Chapman theory.

3.4. The double-layer potentials. The potential differences across the two space-charge regions are expressed by the quantities $\phi_c(0)$ and $\phi_s(0)$. To express these two quantities in terms of the physical constants of the system we note first that according to (23) and (24)

$$\phi_s(0) - \phi_c(0) = \ln(c_A/c_A^0). \quad (50)$$

The second relation between $\phi_s(0)$ and $\phi_c(0)$ is obtained from the condition that the system as a whole is electrically neutral, i.e. from the equation

$$\sigma_c + \sigma_s = 0, \quad (51)$$

so that in conjunction with the expressions for σ_c and σ_s given in §§ 3.2 and 3.3, equations (50) and (51) determine $\phi_c(0)$ and $\phi_s(0)$ uniquely in terms of the ratio (c_A/c_A^0) .

When the general expressions (36) and (48) are used in (51), the resulting equations have to be solved numerically. Explicit formulae for $\phi_c(0)$ and $\phi_s(0)$ may, however,

be derived if the conditions are such that the simplified expressions (37) and (49) may be used in (51). In these circumstances one finds

$$\left. \begin{aligned} \phi_c(0) &= \ln \{1 + (\beta/\gamma)\} - \ln \{1 + \beta\gamma\}, \\ \phi_s(0) &= \ln \{1 + (\gamma/\beta)\} - \ln \{1 + (1/\beta\gamma)\}, \end{aligned} \right\} \quad (52)$$

where
$$\beta = \sqrt{(cWD/\kappa n_0)}, \quad \gamma = \sqrt{(c_A/c_A^0)}. \quad (53)$$

4. FINITE CRYSTALS

4.1. *The equilibrium conditions.* Except when the system is at its isoelectric point, the charge density does not vanish at any point inside a finite crystal, and the equilibrium conditions derived in § 3.1 require some modification.

Consider a plate-like crystal of thickness $2l$, and denote by g_0 and h_0 the concentrations of vacant anion and cation sites at any point in the plane through the centre of the crystal parallel to the surface, i.e. with the origin in this plane and the x -axis perpendicular to it,

$$g_0 = N_B^h(0), \quad h_0 = N_A^h(0). \quad (54)$$

Proceeding exactly as in § 3.1 we may deduce the equation (cf. equation (16))

$$Q_A = -kT \ln \{c_A/(1 - 2pc)\} - kT \ln \{h_0/(N - h_0)\}, \quad (55)$$

except that Q_A is now defined as the work to transfer a cation from the *centre* of the crystal ($x = 0$), across the double layer, and to its hydrated state in solution. If Q_B is defined in a similar manner for anions we also have (cf. equation (17))

$$Q_B = -kT \ln \{c_B/(1 - 2pc)\} - kT \ln \{g_0/(N - g_0)\}. \quad (56)$$

Naturally in all cases of practical interest we shall have

$$\left. \begin{aligned} g_0 &\ll N, \\ h_0 &\ll N, \end{aligned} \right\} \quad (57)$$

so that subtracting (55) and (56) and using the relation $c_A c_B = c_0^2$ we find

$$Q_B - Q_A = 2kT \ln (c_A/c_0) + kT \ln (h_0/g_0), \quad (58)$$

which is the generalization of (20).

Only at the isoelectric point where the double layer of free charges vanishes is the substitution $g_0 = h_0$ strictly valid for any crystal of finite dimensions. As before we denote by W_A and W_B the values of Q_A and Q_B under these conditions, i.e. W_A is the work to transfer a cation from the centre of the crystal to the electrolyte when the system is at its isoelectric point; W_B is similarly defined for anions. c_A^0 , the cation concentration at the isoelectric point, is therefore given by

$$2kT \ln (c_A^0/c_0) = W_B - W_A, \quad (59)$$

which is of the same form as (21). As the dimensions of the crystal are reduced, W_A and W_B must begin to depend upon the size of the crystal, and in these circumstances (59) shows that the ratio (c_A^0/c_0) is also a function of the dimensions of the crystal.

The quantity Λ defined as in (22) is expressed in terms of $V_c(l)$, the potential difference between the surface and the centre of the crystal, and the potential drop in the electrolyte, which we now denote by $V_s(l)$, by the equation

$$\Lambda = V_s(l) - V_c(l). \quad (60)$$

Further, by (58) and (59), we have

$$2|e|\Lambda = 2kT \ln(c_A/c_A^0) + kT \ln(h_0/g_0). \quad (61)$$

In an arbitrary electrolyte the ratio (h_0/g_0) is at present unknown, so that for finite crystals we require two more relations between the three quantities $V_c(l)$, $V_s(l)$ and (h_0/g_0) , to enable these latter to be determined uniquely in terms of the ratio (c_A/c_A^0) . Again these additional relationships are obtained from a consideration of the total charge on the double layer.

4.2. *The charge in the crystal.* We confine our attention to a plate-like crystal of thickness $2l$. With the origin in the plane through the centre of the crystal parallel to the surface and the x -axis perpendicular to it, we denote by $V_c(x)$ the electrical potential at any point x in the crystal. Further, let V_c be defined such that $V_c(x) = 0$ at $x = 0$, then, if $g(x)$ and $h(x)$ are the concentrations of vacant anion and cation sites, it is easy to show that

$$\left. \begin{aligned} h(x) &= [Nh_0 \exp\{eV_c(x)/kT\}]/[N - h_0 + h_0 \exp\{eV_c(x)/kT\}], \\ g(x) &= [Ng_0 \exp\{-eV_c(x)/kT\}]/[N - g_0 + g_0 \exp\{-eV_c(x)/kT\}], \end{aligned} \right\} \quad (62)$$

which are the generalizations of equations (28). If the conditions

$$\left. \begin{aligned} h_0 \exp\{eV_c(l)/kT\} &\ll N, \\ g_0 \exp\{-eV_c(l)/kT\} &\ll N, \end{aligned} \right\} \quad (63)$$

are fulfilled, as they will be if the system is not too far from its isoelectric point, the distribution functions (62) reduce to

$$\left. \begin{aligned} h(x) &= h_0 \exp\{eV_c(x)/kT\}, \\ g(x) &= g_0 \exp\{-eV_c(x)/kT\}. \end{aligned} \right\} \quad (64)$$

The quantities g_0 and h_0 are not independent. The relation between them may be established by considering a process in which a pair of ions are removed from lattice points at the centre of the crystal and placed on the surface in such a way as to build up a new layer of the crystal. Applying the mass-action formula we have

$$g_0 h_0 = n_0^2, \quad (65)$$

where n_0 is constant at constant temperature. n_0 is in fact the concentration of vacant cation or anion sites in the crystal at the isoelectric point when the crystal is uncharged and $g_0 = h_0 = n_0$.

Using the distribution functions (62) in Poisson's equation

$$d^2V_c/dx^2 = -(4\pi e/\kappa)\{g(x) - h(x)\}, \quad (66)$$

the second integration cannot be performed explicitly. However, the expression for σ_c , the charge in the crystal per unit area of surface, requires only the first integral. With the transformations (32), equation (66) integrates under the condition $d\phi_c/dz_c = 0$ at $\phi_c = 0$ to give

$$n_0(d\phi_c/dz_c)^2 = N \ln\{(N - h_0 + h_0 e^{\phi_c})(N - g_0 + g_0 e^{-\phi_c})/N^2\}, \quad (67)$$

so that writing

$$\left. \begin{aligned} 2 \ln \xi(x) &= \phi_c(x), \\ F\{\xi(x); g_0, h_0\} &= \left[\{N - h_0 + h_0 \xi^2(x)\} \{N - g_0 + g_0 \xi^{-2}(x)\} / N^2 \right], \end{aligned} \right\} \quad (68)$$

$$\sigma_c \text{ is given by } \sigma_c = \sqrt{[(\kappa k T N / 2\pi e^2) \ln F\{\xi(l); g_0, h_0\}]}. \quad (69)$$

Finally, we note that integration of (67) yields the equation

$$Nx = \sqrt{(\kappa k T N / 2\pi e^2)} \int_1^\xi [\xi^2 \ln F\{\xi; g_0, h_0\}]^{-\frac{1}{2}} d\xi. \quad (70)$$

When the conditions (63) are satisfied, formulae (67) to (70) reduce considerably. Thus (69) may be written

$$\sigma_c = \sqrt{\left[(2n_0 \kappa k T / \pi e^2) \sinh \{ \phi_c(l) / 2 \} \sinh \left\{ \frac{\phi_c(l) + 2\delta}{2} \right\} \right]}, \quad h_0/g_0 = e^{2\delta}. \quad (71)$$

Furthermore, the potential function may be expressed explicitly in terms of Jacobian elliptic functions. Two cases arise depending on whether g_0 is greater or less than h_0 , i.e. depending on whether the crystal is positively or negatively charged.

(I) $g_0 > h_0$. q , the modulus of the elliptic functions, is given by

$$q = \sqrt{(h_0/g_0)}, \quad (72)$$

and equation (70) degenerates to

$$\xi = \operatorname{cn} bx / \operatorname{dn} bx, \quad b = \sqrt{(2\pi e^2 n_0 / q \kappa k T)}. \quad (73)$$

(II) $g_0 < h_0$. q , the modulus of the elliptic functions, is given by

$$q = \sqrt{(g_0/h_0)}, \quad (74)$$

and equation (70) degenerates to

$$\xi = \operatorname{dn} bx / \operatorname{cn} bx, \quad b = \sqrt{(2\pi e^2 n_0 / q \kappa k T)}. \quad (75)$$

4.3. *The double-layer potentials.* All physical constants of the system being assumed known (in particular we require n_0 , the concentration of vacant sites in an uncharged crystal), the equations of § 3 determine $\phi_c(l)$ as a function of (h_0/g_0) for a given value of l . We may therefore calculate according to (69) σ_c as a function of (h_0/g_0) . The final step is to express (h_0/g_0) in terms of the cation concentration in the electrolyte (or in terms of (c_A/c_A^0)). This is accomplished with the aid of the equation (cf. (60) and (61)),

$$\phi_s(l) - \phi_c(l) = \ln (c_A/c_A^0) + \frac{1}{2} \ln (h_0/g_0),$$

and the condition for electrical neutrality, $\sigma_s + \sigma_c = 0$. In this way the three quantities $\phi_c(l)$, $\phi_s(l)$ and (h_0/g_0) may be calculated as functions of (c_A/c_A^0) for crystals of arbitrary thickness, and a complete description of the charge distribution between the crystal and the electrolyte is obtained.

5. FRENKEL DEFECTS

For crystals containing vacant cation sites and interstitial cations, the analysis of §§ 3 and 4 may be developed along similar lines. The equations are slightly more complicated, but they simplify somewhat if we confine ourselves to dilute solutions

so that the 'activity coefficients' $(1 - 2pc)$ in (8) may be set equal to unity.† In these circumstances the isoelectric point is given by

$$-2kT \ln c_A^0 = W_A^i + W_A, \quad (76)$$

where W_A^i is the work to transfer a cation from an interstitial position in the crystal to the electrolyte when there is no double layer of free charges present. The quantity Λ defined by

$$\Lambda = V_s(l) - V_c(l),$$

is given by

$$2|e|\Lambda = 2kT \ln(c_A/c_A^0) + kT \ln(\alpha h_0/g_0),$$

where α is the ratio of the number of interstitial positions to the number of ion pairs ($\alpha = 2$ for simple cubic structures), and we have used the notation (cf. equation (54))

$$g_0 = N_A^i(0), \quad h_0 = N_A^h(0).$$

Equations (64) to (75) remain unaltered if the function F in (68) to (70) is redefined according to the equation

$$F\{\xi(x); g_0, h_0\} = [\{N - h_0 + h_0 \xi^2(x)\} \{\alpha N - g_0 + g_0 \xi^{-2}(x)\} / \alpha N^2].$$

6. DISCUSSION OF THE METHOD

In this section we discuss briefly the approximations involved in the theory of the double layer developed in §§ 2 to 5. The most important points which arise concern the validity of the distribution functions (42) and (62). Consider first the distribution of vacant sites in the crystal.

As we have already mentioned, interactions between vacant sites have been neglected in deriving (62). The influence of these interactions clearly becomes more important as the total concentration of vacant sites increases, so that unless

$$\begin{aligned} h_0 \exp\{eV_c(l)/kT\} &\ll N, \\ g_0 \exp\{-eV_c(l)/kT\} &\ll N, \end{aligned}$$

the distribution functions (62) are only very crude approximations to the true state of affairs in the space-charge region in the crystal. The possibility of improving on (62) by specifically introducing defect interactions lies beyond the scope of the present work, but I do not believe that such refinements would seriously affect the general features of the charge distribution between the crystal and the electrolyte.

Apart from the considerations of the last paragraph, there is another effect which is neglected in the distribution functions (62). It is that any defect located near the interface comes under the influence of additional forces arising essentially from the fact that the crystal does not extend to infinity in all directions. The nature of these 'adsorption' forces can be seen by considering the crystal and the electrolyte as continuous media of dielectric constants κ and D . For most ionic crystals $\kappa \sim 10$, whilst $D \sim 80$, so that the image force is directed towards the interface. The potential energy of these adsorption forces should be included in the exponentials in (62). Of

† Of course, owing to our neglect of ionic interactions, the analysis for Schottky defects is really applicable only when $2pc \ll 1$.

course these forces may be ignored in the first approximation, but I believe that, in a more detailed theory, their influence on the distribution functions (62) is one of the first points which would require deeper investigation.

The remarks of the foregoing paragraphs may be applied *mutatis mutandis* to the functions (42) governing the distribution of ions in the electrolyte. However, in view of the important part played by adsorption forces in determining the structure of the double layer for many solid/electrolyte systems, some further comments on their importance in the present theory seem to be called for. We note first that the image force is directed *away* from the interface. This is not, however, the only force acting on an ion in solution near the interface. There are, in addition, three other terms: (a) electrostatic, (b) van der Waals attraction, (c) short-range repulsion. (a) and (b) have been discussed by Lennard-Jones & Dent (1928) and (except for $\langle 111 \rangle$ surfaces) both can be given analytic forms. The repulsive energy (c) setting in at very close distances of approach can be represented fairly well as an exponential function of the distance between ionic centres (Bleich & Mayer 1934). For simple monovalent ions at distances greater than the lattice parameter outside a $\langle 100 \rangle$ surface of a crystal having the NaCl structure, the potential energy of the forces (a) to (c) does not amount to more than a few hundredths of a volt. The smallness of this potential energy, coupled with the fact that the image force is directed away from the interface (the potential energy of this force is closely connected with the process of hydration), indicates that simple monovalent ions can scarcely be adsorbed on to $\langle 100 \rangle$ surfaces. On the other hand, for large, weakly hydrated ions, the increased importance of the van der Waals term (b), coupled with the small amount of work performed against the hydration forces, may clearly result in the adsorption of such ions. The presence in the electrolyte of 'surface-active' ions of this sort in high enough concentrations so that the adsorbed charge is comparable with the space charge in the crystal, clearly invalidates the theory of the double layer developed in §§ 2 to 5.

7. THE SYSTEM SILVER BROMIDE/ELECTROLYTE

7.1. *The potential drop in the crystal and the charge on the double layer.* In this section we apply the analysis of §§ 4 and 5 to the particular case of silver bromide in contact with an electrolyte. The results of a preliminary attempt to discuss this system quantitatively have already been published (Grimley & Mott 1947). In this work we proceeded on the assumption that the lattice defects normally present in the crystal are of the Frenkel type, and attempted to estimate the energies W_A^i and W_A appearing in (76) by direct calculation. For this purpose we may write

$$\begin{aligned} W_A^i &= U_A^i - H_A + |e| \chi, \\ W_A &= U_A - H_A + |e| \chi, \end{aligned} \quad (77)$$

where U_A^i is the interaction energy between an interstitial ion and its surroundings, U_A is similarly defined for an ion at its normal lattice site and H_A is the corresponding quantity for the ion in solution, i.e. H_A is the energy of hydration. Finally, χ is the 'phase-boundary potential' (nomenclature of Lange 1933) at the isoelectric point.

If P is the dipole moment per unit area of the interface in the direction of the outward normal to the crystal surface χ is given by

$$\chi = 4\pi P.$$

We calculated W_A^i and W_A from (77) by setting $\chi = 0$, using Verwey's (1940) estimate of the hydration energy of silver ions and calculating U_A^i and U_A by the method first employed by Mott & Littleton (1938) in their work on the alkali halides. However, an investigation into the principles underlying these calculations (to be published) has revealed that our original estimates of U_A^i and U_A may be subject to serious error. The error arises from our neglect of the homopolar character of the binding in silver bromide; for the alkali halides the error would be unimportant. Furthermore, it must be admitted that the substitution $\chi = 0$ is an arbitrary procedure, we have no knowledge of either the magnitude or sign of this quantity.

In the light of these considerations I have abandoned the attempt to fix the isoelectric point theoretically, and in this work I shall use the experimentally determined value. Actually the situation in this respect is not entirely satisfactory, and widely divergent results are to be found in the literature (see, for example, Julien 1933; Reyerson, Kolthoff & Coad 1947). However, Jonker (1943), working with silver bromide sols, has obtained reproducible results† which indicate that the isoelectric point of a freshly prepared sol lies between $p_{Ag} = 5$ and $p_{Ag} = 6$. The subsequent calculations of this paper are based on the existence of the isoelectric point at $p_{Ag} = 5.8$. In the concentration units here employed

$$c_A^0 = 2.3 \times 10^{-8}. \quad (78)$$

For the concentration of defects in silver bromide at room temperature I take the value of 10^{17} as deduced from the conductivity data of Koch & Wagner (1937) (cf. Mott & Gurney 1940, p. 47), but following Mitchell (1949) I shall discuss the charge in the crystal in terms of Schottky rather than Frenkel defects. In accordance with (65) therefore

$$g_0 h_0 = n_0^2 = 10^{34}. \quad (79)$$

In this connexion it should be pointed out that having fixed the isoelectric point empirically, the distinction between the two types of defects is rather trivial in the present analysis. Of course, physical properties of the solid which are influenced by the presence of the space charge inside it, are vastly different in the two cases (cf. § 7.2).

With the information contained in (78) and (79) I calculated, according to the method outlined in § 4.3, $V_c(l)$ and $V_s(l)$ as functions of the silver-ion concentration for crystals of different thicknesses.‡ The dependence of these potentials on the crystal thickness ($2l$) sets in sharply at $l < 10^{-6}$ cm. For $l > 10^{-6}$ cm. the dependence is

† The isoelectric point drifts to higher silver ion concentrations as the sol ages. The reason for this behaviour (as well as the divergent results of other workers) is not clear, but it seems to indicate that the nitrate ion is surface active in the system silver bromide/electrolyte (cf. § 6).

‡ Over the greater part of the concentration range (roughly $10.6 > p_{Ag} > 1.2$) the distribution functions (42) and (62) reduce to the Boltzmann laws (44) and (64).

negligible and, for the purposes of calculating $V_c(l)$ and $V_s(l)$, the crystal may be treated as infinite and the equations of § 3 used. The results are shown in figure 1.

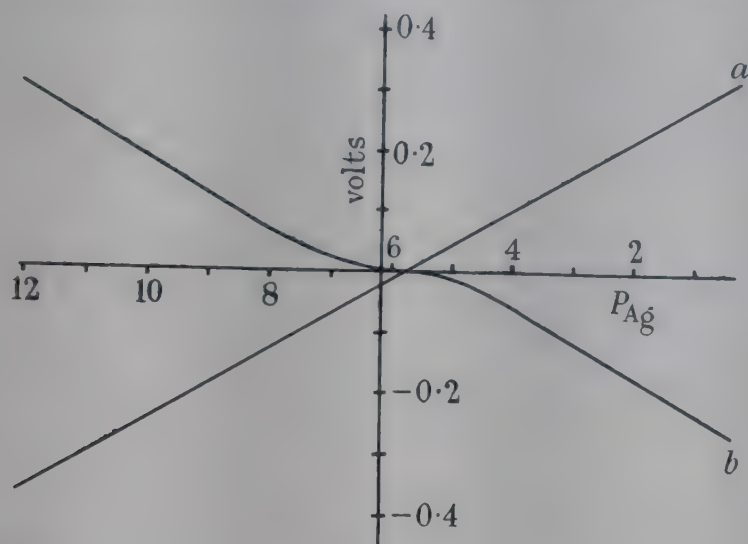


FIGURE 1. Double-layer potentials as functions of the silver-ion concentration for silver bromide in contact with an electrolyte: a , Λ ; b , $V_c(l)$. The sum $(a+b)$ is the potential drop in the electrolyte. The electrolyte contains either silver salts or soluble bromides; there is no indifferent electrolyte present.

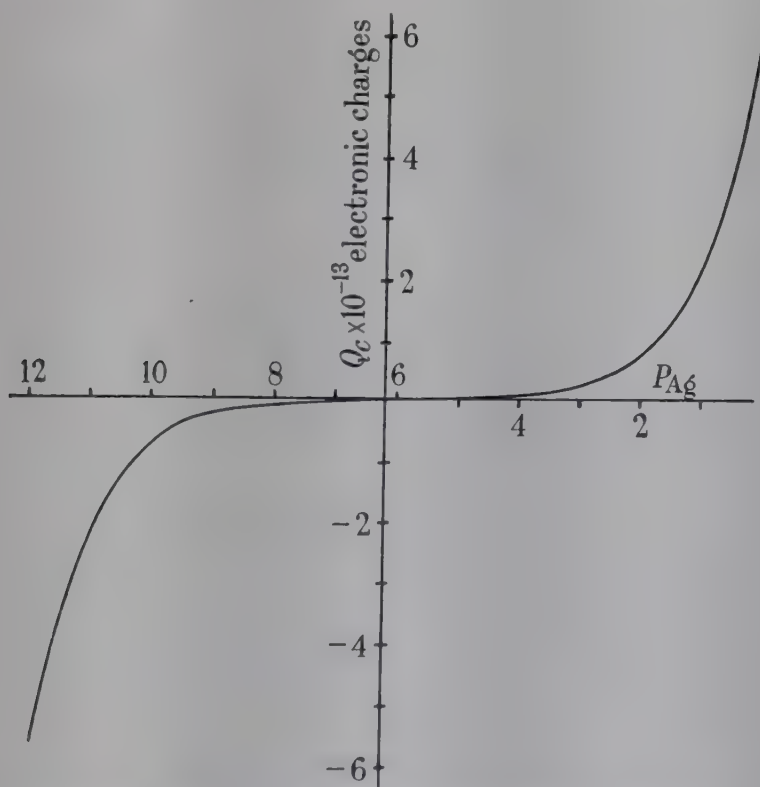


FIGURE 2. The charge per unit area of a silver bromide crystal immersed in an electrolyte. (No indifferent electrolyte present.)

The charge in the crystal per unit area of surface (for $l > 10^{-6}$ cm.) is readily calculated from (37) and the results are shown in figure 2. Now the fundamental point of the present theory is that this charge arises through the presence, in the crystal, of vacant cation and anion sites in unequal concentrations. To be more precise, on the

silver side of the isoelectric point, vacant anion sites are in excess, whilst on the bromide side the reverse is true. The origin of this charge is illustrated most clearly by plotting the concentrations of vacant cation and anion sites across the crystal. I give the results for an electrolyte with $p_{Ag} = 2$. In this electrolyte $h(x)$ the concentration of vacant cation sites at distance x from the centre of the crystal is given by (cf. equations (64) and (74))

$$h(x) = qn_0 \operatorname{cn}^2 bx / \operatorname{dn}^2 bx, \quad (80)$$

where $b = \sqrt{(2\pi e^2 n_0 / q \kappa k T)}$ and q , the modulus of the elliptic functions, is given by $q = \sqrt{(h_0 / g_0)}$. h_0 and g_0 , it will be recalled, are the concentrations of vacant cation and anion sites at the centre of the crystal ($x = 0$). For $g(x)$, the concentration of vacant anion sites, we shall have

$$g(x) = (n_0 / q) \operatorname{dn}^2 bx / \operatorname{cn}^2 bx. \quad (81)$$

Defect concentrations calculated from (80) and (81) for two thicknesses of crystal are shown in figure 3. Actually curves of this sort are quite easily obtained for crystals where $l > 10^{-6}$ cm. now that we are in possession of the results shown in figure 1, for the equations (cf. equations (73) and (75))

$$\exp \{eV_c(l) / kT\} = \operatorname{dn}^2 bl / \operatorname{cn}^2 bl \quad (c_A > c_A^0),$$

$$\exp \{eV_c(l) / kT\} = \operatorname{cn}^2 bl / \operatorname{dn}^2 bl \quad (c_A < c_A^0),$$

may be used to determine q as a function of c_A for given l .

7.2. *Self-diffusion in the solid.* Inasmuch as the self-diffusion coefficients of the ions are proportional to the concentrations of vacant sites, the results of § 6 make it clear that the rates of diffusion of both silver and bromide ions in the crystal will depend, in general, upon the dimensions of the crystal and the p_{Ag} of the electrolyte. If, for example, the self-diffusion coefficients are calculated, using ordinary diffusion theory, from the observed rates of exchange of radioactive ions between the electrolyte and the crystal, the self-diffusion coefficient of cations will be enhanced on the bromide side and diminished on the silver side of the isoelectric point. For anions the reverse will hold. Only at the isoelectric point could the true self-diffusion coefficients be calculated in this way. Whilst with sufficiently large crystals no difficulties of this sort will occur, ionic exchange experiments of the type described above have already been carried out using photographic plates (Zimens 1946) and silver bromide suspensions (Langer 1943). In these experiments the particle sizes lie in the range 10^{-5} to 10^{-4} cm., and inspection of figure 3 shows that the presence of the space-charge region of lattice defects will be felt, at least at the lower end of the range of particle size. In this section I shall derive approximate relationships between the true self-diffusion coefficients, i.e. those for the material in bulk, and the apparent self-diffusion coefficients calculated from the results of radioactive ion exchange experiments using ordinary diffusion theory. A complete discussion of this problem presents considerable difficulty and, for reasons which I shall enumerate later, can hardly be justified at the present time.

To fix our ideas, consider a plate-like crystal of silver bromide immersed in an electrolyte containing radioactive silver ions. If $n^*(x, t)$ is the concentration of

radioactive silver ions in the crystal at any point x and at time t , then, in the absence of any space-charge region in the crystal, the differential equation for n^* is

$$D(\partial^2 n^* / \partial x^2) = \partial n^* / \partial t, \quad (82)$$

D being the self-diffusion coefficient of silver ions. If $2l$ is the width of the crystal and n_0^* the concentration of radioactive silver ions at the surface, we take the boundary conditions on (82) in the form†

$$\left. \begin{aligned} \partial n^* / \partial x &= 0 & \text{at } x &= 0, \\ n^* &= n_0^* & \text{at } x &= \pm l, \\ n^* &= 0 & \text{at } t &= 0 \quad (-l < x < l). \end{aligned} \right\} \quad (83)$$

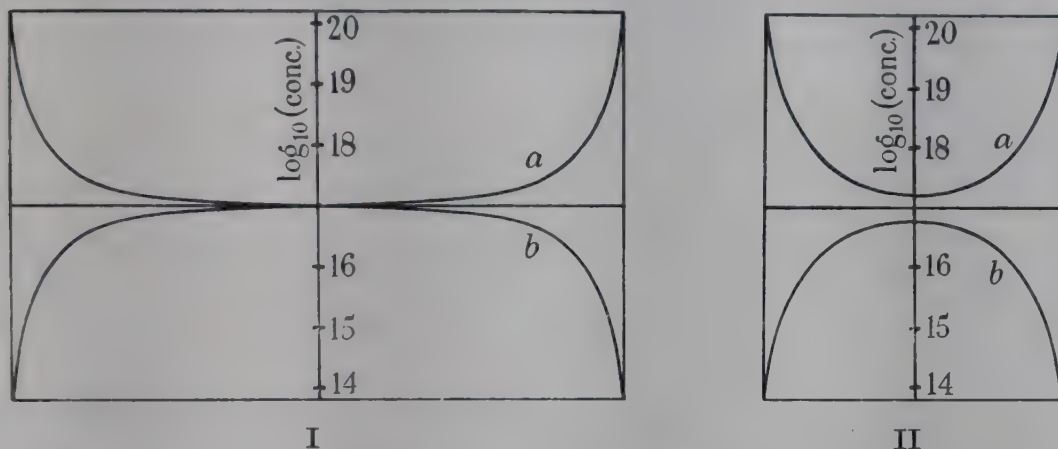


FIGURE 3. Defect concentrations in plate-like crystals of silver bromide immersed in an electrolyte with $p_{Ag} = 2$. Crystal thickness, I, 110 $m\mu$; II, 55 $m\mu$. The upper curves (a) are vacant anion sites, the lower curves (b) are vacant cation sites. (No indifferent electrolyte present.)

Under these conditions τ_0 , the half-life of the diffusion process, is given by (see, for example, Carslaw & Jaeger 1941, p. 94)

$$\tau_0 = 4l^2 / \pi^2 D. \quad (84)$$

Now the theory of this paper has shown that only at the isoelectric point are the concentrations of defects uniform across the thickness of the crystal. In any other electrolyte the concentrations of vacant cation and anion sites are given in terms of Jacobian elliptic functions by equations of the form (80) and (81). (The deviations from these laws which occur outside the range $10.6 > p_{Ag} > 1.2$ are unimportant in the present discussion.) Equation (82) is now replaced by

$$D_A^h \frac{\partial}{\partial x} \left\{ (N_A^h / N) \frac{\partial n^*}{\partial x} \right\} = \frac{\partial n^*}{\partial t}, \quad (85)$$

where D_A^h is the diffusion coefficient of cation vacancies. On the bromide side of the isoelectric point N_A^h is given by

$$N_A^h(x) = (n_0/q) \operatorname{dn}^2 bx / \operatorname{cn}^2 bx,$$

† These boundary conditions are not strictly appropriate to the experiments of Zimens & Langer. In these experiments n_0^* is a function of time. This is an additional complication which, however, I do not intend to discuss.

with $b = \sqrt{(2\pi e^2 n_0 q \kappa k T)}$. Introducing this value into (85) and making the transformation $bx = y$, there results

$$\frac{\partial}{\partial y} \left\{ \frac{dn^2 y}{cn^2 y} \frac{\partial n^*}{\partial y} \right\} = \left\{ \frac{qN}{b^2 n_0 D_A} \right\} \frac{\partial n^*}{\partial t}. \quad (86)$$

Under the substitution $n^*(y, t) = Y(y) T(t)$, (86) separates to give

$$n^*(y, t) = Y(y) \exp \{ -\lambda b^2 n_0 D_A t / qN \},$$

where λ is an eigenvalue of the equation

$$\frac{\partial}{\partial y} \left\{ \frac{dn^2 y}{cn^2 y} \frac{\partial Y}{\partial y} \right\} + \lambda Y = 0. \quad (87)$$

Denote the lowest non-vanishing eigenvalue of (87) by λ_1 then τ , the half-life of the exchange, is given by

$$\tau = (qN / b^2 n_0 D_A \lambda_1).$$

Defining an apparent self-diffusion coefficient by means of the relation $\tau = 4l^2 / \pi^2 D_{app}$, we readily find

$$D_{app} / D = (8\pi e^2 n_0 / \kappa k T) (\lambda_1 l^2 / \pi^2 q). \quad (88)$$

Approximations to λ_1 are not difficult to obtain. Thus, on making the transformations (cf. Courant & Hilbert 1924, p. 335)

$$z = Y \sqrt{(dn y / cn y)} \quad \text{and} \quad m = \int_0^y (cn y / dn y) dy,$$

$$\text{equation (87) becomes} \quad d^2 z / dm^2 - f(m) z + \lambda z = 0, \quad (89)$$

where, on writing $p^2 = (1/q^2)(q^2 - \tanh^2 qm)$, $f(m)$ is given by

$$p^4 f(m) = 5q'^4 + p^2(4q'^2 - 7q'^4) - 2p^4(q'^2 - q'^4).$$

Here q' is the complementary modulus of the elliptic functions. Now the eigenvalues of (89) differ from the eigenvalues of the equation

$$d^2 z / dm^2 + \lambda z = 0 \quad (90)$$

by no more than Δ , a finite quantity given by

$$\Delta = \int_0^\mu z^2 f(m) dm,$$

where

$$\mu q = \tanh^{-1} \{ q \operatorname{sn}(bl/K) \}, \quad (91)$$

and z is an eigenfunction of (90). In (91), K is the complete elliptic integral of the first kind. The eigenvalues of (90) appropriate to our problem are $0, \pi^2/4\mu^2, 9\pi^2/4\mu^2$, etc., so that our zero-order approximation to λ_1 is

$$\lambda_1^0 = \pi^2/4\mu^2.$$

As our first-order approximation we take

$$\lambda_1^1 = \lambda_1^0 + \int_0^\mu f(m) \cos^2(\pi m/2\mu) dm,$$

and the true value of λ_1 lies between these two. Calculation shows that the results in the first approximation are some 20 % greater than those in the zero order. Since our main interest in this problem is to estimate orders of magnitude, the degree of accuracy furnished by the zero-order approximation is clearly sufficient for our purpose. The results of this calculation for crystals of several thicknesses are shown in figure 4. The results on the silver side of the isoelectric point were obtained in precisely the same way. In excess silver the ratio $D_{\text{app.}}/D$ is of course less than unity, and I have exhibited the results by plotting the inverse ratio $D/D_{\text{app.}}$.

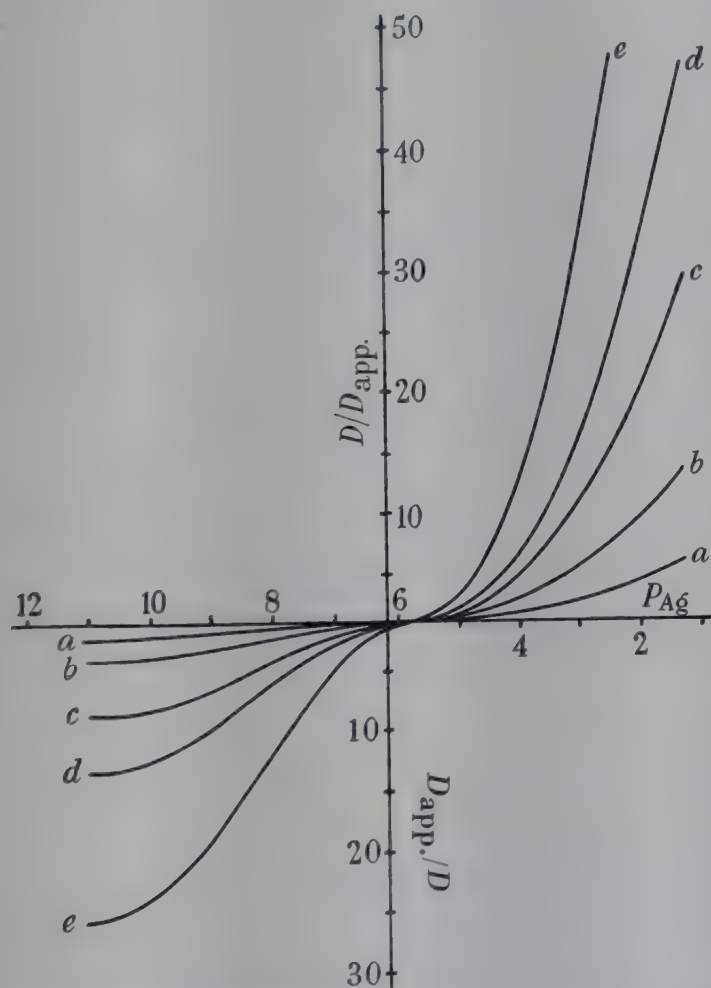


FIGURE 4. Apparent self-diffusion coefficients of silver ions in plate-like crystals of silver bromide immersed in an electrolyte. Crystal thickness: *a*, 110 $m\mu$; *b*, 55 $m\mu$; *c*, 34 $m\mu$; *d*, 27 $m\mu$; *e*, 20 $m\mu$. (No indifferent electrolyte present.)

As we have already mentioned, a certain amount of experimental work on the exchange of radioactive ions between solid silver bromide and an electrolyte has already been carried out. Owing to the low solubility product of the salt, silver-ion exchange is of necessity observed on the silver side of the isoelectric point; similarly, bromide-ion exchange can only be observed in the presence of excess bromine. Our theoretical results show therefore that self-diffusion coefficients, calculated from

these experiments by straightforward diffusion theory, will be less than the corresponding quantities for the material in bulk, the difference becoming increasingly important as the dimensions of the crystals are reduced beyond $\sim 2 \times 10^{-5}$ cm.† There is, however, a difficulty which we have so far ignored, namely, that the rate of exchange may be governed by a surface reaction rather than by diffusion in the solid. If this is actually the case, the effect which we have discussed above may be completely masked. Zimens (1946) concluded that for photographic plates (particle size 10^{-4} cm.) the rate-determining step is a surface reaction. The system photographic plate/electrolyte is, however, more complicated than the ideal simple case which is the subject of the present work, and it would be optimistic to expect any of the theory here developed to be applicable, as it stands, to Zimens's work.

The work of Langer (1943) on aqueous suspensions of silver bromide is of considerably more interest from the point of view of the present theory. This author came to three main conclusions: (a) both silver and bromide ions exchange with the precipitate; (b) the exchange goes to completion, i.e. it is not confined to the surface layers of the solid; (c) diffusion in the solid is the rate-determining step. Now although the necessary calculations were not carried out by Langer, sufficient information is actually contained in the original paper to enable the apparent self-diffusion coefficients and the dimensions of the particles to be estimated. The particles being assumed spherical, the radii can be calculated from the recorded data on the adsorption of methylene blue (surface area per molecule $\sim 3 \times 10^{-15}$ cm.²). The results for silver-ion exchange with five different precipitates are shown in table 1. In these calculations a value of $D = 10^{-13}$ cm.²/sec., as derived from the data

TABLE 1. APPARENT SELF-DIFFUSION COEFFICIENTS FOR SILVER IONS
CALCULATED FROM LANGER'S DATA

$D/D_{\text{app.}}$	radius in $m\mu$
25	73
22	100
15	130
12	290
11	610

of Koch & Wagner (1937), has been used. The results are at once significant. The consistent trend in $D/D_{\text{app.}}$ with particle size is precisely the behaviour to be expected on the basis of the theory developed earlier in this section. As already pointed out, these exchange experiments are subject to boundary conditions which are slightly different from those used in our theoretical discussion. For this reason, and also owing to the different geometry, a comparison of the results of table 1 with our theoretical calculations is not strictly permissible. Further, the theoretical results are uncertain in so far as the equilibrium concentration of defects in the solid is not known to within a power of 10, whilst the experimental results of table 1 depend upon the estimation of particle size from dye adsorption data. In these circumstances

† In line with all the calculations of the present paper, this figure is based on a value of 10^{17} for the equilibrium concentration of lattice defects. Had we taken 10^{16} defects per unit volume, this dimension would be increased to $\sim 6 \times 10^{-5}$ cm.

I do not feel justified in attempting a more detailed discussion of these ionic exchange experiments at the present time. There seems little doubt, however, that, provided one is willing to tamper with the value used for the equilibrium concentration of lattice defects (see the footnote on p. 60), the experimental results can be accounted for on the basis of the theory developed above.

Much of the work here reported was carried out at the H. H. Wills Physical Laboratory, University of Bristol, and I should like to express my appreciation to Professor N. F. Mott, F.R.S., for his sustained interest in this research. I am also grateful to Dr J. W. Mitchell for many valuable discussions concerning the application of the theory to the system silver bromide/electrolyte.

REFERENCES

- Bleich, W. E. & Mayer, J. E. 1934 *J. Chem. Phys.* **2**, 252.
 Carslaw, H. S. & Jaeger, J. C. 1941 *Operational methods in applied mathematics*. Oxford University Press.
 Chapman, D. L. 1913 *Phil. Mag.* **25**, 475.
 Courant, R. & Hilbert, D. 1924 *Methoden der Mathematischen Physik*, **1**. Berlin: J. Springer.
 Fowler, R. H. 1932 *Proc. Roy. Soc. A*, **136**, 391.
 Fowler, R. H. & Guggenheim, E. A. 1939 *Statistical thermodynamics*. Cambridge University Press.
 Frenkel, J. 1926 *Z. Phys.* **35**, 652.
 Gouy, G. 1910 *J. Phys.* **9**, 457.
 Grimley, T. B. & Mott, N. F. 1947 *Disc. Faraday Soc.* **1**, 3.
 Guggenheim, E. A. 1929 *J. Phys. Chem.* **33**, 842.
 Jonker, G. H. 1943 Dissertation, Utrecht.
 Jost, W. 1933 *J. Chem. Phys.* **1**, 466.
 Julien, P. F. J. 1933 Dissertation, Utrecht.
 Koch, E. & Wagner, C. 1937 *Z. phys. Chem. B*, **38**, 295.
 Kruyt, H. R. & van der Willigen, P. C. 1928 *Z. phys. Chem.* **139**, 53.
 Lange, E. 1933 *Handb. Exp. Phys.* **xii/2**, 265.
 Langer, A. 1943 *J. Chem. Phys.* **11**, 11.
 Lennard-Jones, J. E. & Dent, B. M. 1928 *Trans. Faraday Soc.* **24**, 92.
 Mitchell, J. W. 1949 *Phil. Mag.* **40**, 249.
 Mott, N. F. & Gurney, R. W. 1940 *Electronic processes in ionic crystals*. Oxford: Clarendon Press.
 Mott, N. F. & Littleton, M. J. 1938 *Trans. Faraday Soc.* **34**, 485.
 Reyerson, L. H., Kolthoff, I. M. & Coad, K. 1947 *J. Phys. Colloid Chem.* **51**, 321.
 Stern, O. 1924 *Z. Elektrochem.* **30**, 508.
 Verwey, E. J. W. 1937 *Hydrophobic colloids*, p. 58. Amsterdam: D. B. Centen's Uitg. Mij. N.V.
 Verwey, E. J. W. 1940 *Chem. Weekbl.* **36**, 530.
 Verwey, E. J. W. & Kruyt, H. R. 1933 *Z. phys. Chem. A*, **167**, 137.
 Verwey, E. J. W. & Overbeek, J. Th. G. 1948 *Theory of the stability of lyophobic colloids*. Amsterdam: Elsevier.
 Wagner, C. & Schottky, W. 1930 *Z. phys. Chem. B*, **11**, 163.
 Zimons, K. E. 1946 *Ark. Kemi Min. Geol. A*, **23**, 16.

The crystal structure of horse met-myoglobin

I. General features: the arrangement of the polypeptide chains

BY J. C. KENDREW

Cavendish Laboratory and Molteno Institute, University of Cambridge

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 6 July 1949—
Revised 12 October 1949)

[Plate 1]

The preliminary results of an X-ray study of crystalline horse met-myoglobin are described. This protein crystallizes in the monoclinic space-group $P2_1$ with two molecules in general positions in the unit cell. The optical properties of the crystals show that the haem groups are parallel and lie approximately in the plane containing b and the internal bisector of β .

a and c Patterson projections contain ridges of high vector density perpendicular to b and $15.4 \text{ \AA}$ apart; it is deduced that in real space the contents of the cell are arranged in pronounced layers of alternately high and low electron density, also perpendicular to b and $15.4 \text{ \AA}$ apart. The b projections also exhibit rods of high vector density, $9.5 \text{ \AA}$ apart and inclined at 20° to a . These rods are identified as the vector equivalents of the polypeptide chains, which thus run parallel to $[20\bar{1}]$ and $9.5 \text{ \AA}$ apart. A Patterson projection along $[20\bar{1}]$ shows the rods well-resolved and in end-on view; by making certain plausible assumptions a Fourier projection has been made along the same direction.

During shrinkage the cell dimensions change in such a way that the inter-chain distance alters very little; it is deduced that the molecules have flat sides parallel to (102) and that the liquid of crystallization is situated mainly on their $[20\bar{1}]$ faces. During shrinkage liquid is withdrawn and the molecules slide past one another along their flat sides. It is not possible to determine the whole shape of the molecule unequivocally; the most plausible model is a platelet built of four parallel co-planar lengths of chain each $54 \text{ \AA}$ long (see figure 13).

There are striking analogies between the structures of myoglobin and of horse haemoglobin. The type of chain folding seems to be the same in the two proteins, and probably the same as in α -keratin; and it may be that myoglobin has a structure analogous to that of one of the four layers of which haemoglobin is believed to be built up, the molecular weights of the two proteins being in the ratio 1:4.

1. INTRODUCTION

Myoglobin appeared to be a suitable protein for X-ray analysis because of its low molecular weight, because it is so closely related in properties and functions to haemoglobin (whose crystal structure has been more intensively studied than that of any other protein), and because almost complete amino-acid analyses are now available for two species (horse and whale). On the other hand, large single crystals of myoglobin are difficult to prepare from any species so far investigated. Among the largest yet obtained are those of horse myoglobin, used in the present research, but even these are far from satisfactory.

Myoglobin is a conjugated protein occurring in muscle cells. Its prosthetic group, ferroprothaem, is identical with that of haemoglobin, and like haemoglobin it forms oxy-, carboxy- and met-derivatives. Its function seems to be that of oxygen store and intermediary between the blood haemoglobin and the cellular oxidation-reduction systems, but the details are still obscure. Earlier research on myoglobin

has been reviewed by Millikan (1939), and further information is given by Vieil (1942). Here it will suffice to mention a few facts bearing directly on structural problems.*

(a) The molecular weight is in the neighbourhood of 17,000 (Svedberg 1939; Roche & Vieil 1940); that is to say just one-quarter of the molecular weight of haemoglobin (67,000 to 68,000).

(b) The iron content is 0.345, expressed as a percentage of dry protein weight (Theorell 1932). This is identical with the corresponding figure for haemoglobin, which contains four iron atoms per molecule; hence the myoglobin molecule must contain one iron atom and one haem group.

(c) The oxygen dissociation curve is strictly a rectangular hyperbola (see Millikan 1939). This would be expected for a molecule containing only one haem group, on the basis of accepted theory which attributes deviations from hyperbolic shape to interactions between several haem groups in the same molecule.

(d) Some data on the amino-acid composition of myoglobin have been given by Rossi-Fanelli (1940, 1949) and by Roche, Derrien & Vieil (1942), but the first almost complete analysis is that of Tristram (1949). He shows that the composition of myoglobin bears no obvious resemblance to that of haemoglobin, so there is no basis for the suggestion that myoglobin is in any sense a precursor of the larger protein. However, in spite of the difference in amino-acid composition, the two proteins have very similar iso-electric points (myoglobin, 6.99; haemoglobin, 6.78; see Theorell 1932). Porter & Sanger (1948), using the 'end-group' method, have found that myoglobin contains twenty lysine ϵ -NH₂ groups and one α -NH₂ group, belonging to glycine; they deduce that, apart from possibilities of cyclic polypeptides, the molecule is one single polypeptide chain, built up of about 146 amino-acid residues (Tristram 1949).

(e) The shape of the myoglobin molecule in solution has been discussed by Wyman & Ingalls (1943; see also Wyman 1948) on the basis of measurements of quantities such as sedimentation and diffusion constants, viscosity, and dielectric constant. These authors reach no final conclusion, partly because the hydration of myoglobin in solution is unknown; but they conclude, as a compromise between all the data, that if the equivalent ellipsoid is oblate (as they consider most probable) its asymmetry ratio is 3.6:1; this result implies an abnormally low value for the hydration (0.1 g. H₂O/g. protein). On the other hand, Gutfreund (1950), taking into account only the sedimentation and diffusion constants, finds that the maximum possible asymmetry is about 2:1 for a completely unhydrated molecule, and that even if the shape were taken as spherical the hydration could not exceed 0.15 g. H₂O/g. protein. These discrepant conclusions suggest that the matter requires further investigation.

The following two sections, §§ 2 and 3, give details of the experimental methods used in this research, and of the crystallographic data for crystals of horse met-myoglobin. § 4 gives an account of the deductions which may be drawn from the diffraction pattern of the wet crystals by plotting Patterson projections; § 5 describes the main features of the diffraction pattern of the dry crystal. Finally,

* Throughout this paper the term myoglobin denotes that derived from horse heart, unless otherwise specifically stated.

§6 contains a discussion of the conclusions which may be drawn about the structure and, in particular, about the arrangement of polypeptide chains in the molecule. These conclusions are summarized in §7.

An account of the earlier stages of this research has already been published (Kendrew 1949).

2. EXPERIMENTAL METHODS

Preparation of crystals

It has been known for many years that crystalline myoglobin can be obtained by salting out its aqueous solution with ammonium sulphate; the crystals are long, fine needles. It was found, however, that when prepared by this method the needles were invariably too thin for X-ray single-crystal work. The crystals actually used in the research were prepared by Mr M. W. Rees of the Department of Biochemistry, Cambridge University, who discovered that myoglobin could be crystallized from concentrated phosphate buffer instead of from ammonium sulphate. For crystallization both myoglobin and buffer must be in high concentration—the latter between 3 and 4 molar; mixtures of $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$ and K_2HPO_4 were used, being the most soluble pair of acid and alkaline phosphates available, and the pH being about 6.4 (details of an essentially similar crystallization of whale myoglobin have been published by Keilin & Schmid 1948).

The crystals so prepared are much more suitable for X-ray purposes than those obtained from ammonium sulphate solution; the needles are flattened out into plates, often several millimetres long and 1 mm. across. Even so they are very thin, and it was necessary to use extremely long X-ray exposures.

Technique of X-ray analysis

Standard methods of mounting the wet crystals were used. All photographs were taken with Cu $K\alpha$ radiation. After a few preliminary pictures had been taken with a conventional single-crystal goniometer, nearly all the rest of the work was done with the Buerger Precession Camera. A full account of this apparatus has been given by its inventor (Buerger 1944); its special advantage is that, if the X-ray beam is directed along any zone axis of the crystal, it gives an undistorted representation of any desired reciprocal lattice plane perpendicular to this axis, up to a Bragg angle of 35° . It is thus particularly applicable to the study of protein crystals, which do not reflect at Bragg angles greater than about 18° or 20° , so that a single photograph extends over the whole of the reflecting region of any reciprocal plane passing through the origin. The only limitation is that for all reciprocal planes other than those passing through the origin a central region round the zone-axis is missing; however, this is immaterial when only two-dimensional projections are to be carried out (as in the present research), since only the basal reciprocal planes need then be recorded.

As an example of an X-ray photograph taken with the Buerger camera, figure 1 (plate 1) shows the *ab* plane of horse myoglobin.

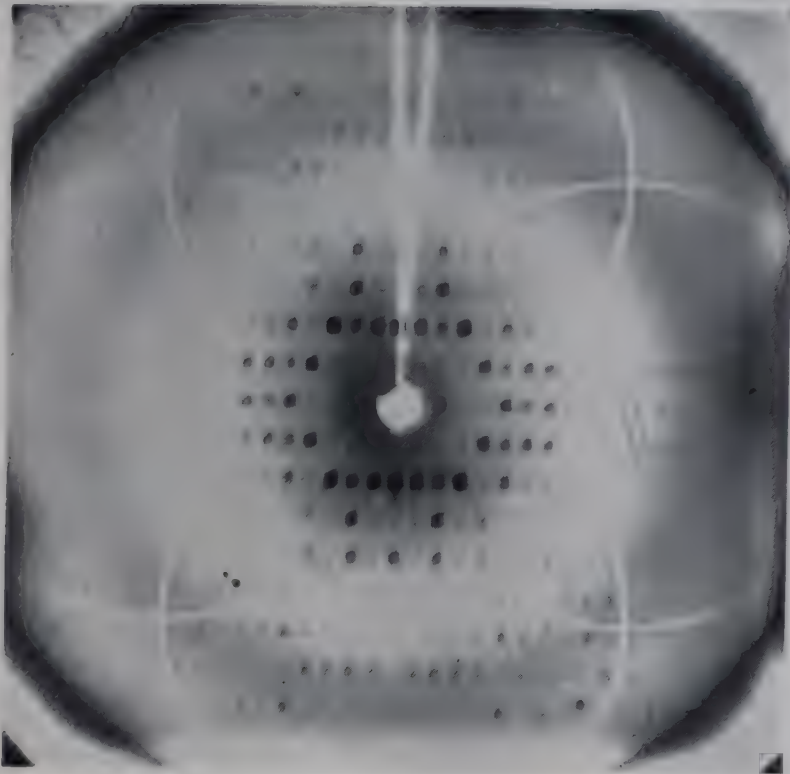


FIGURE 1. X-ray diffraction photograph of the ab reciprocal lattice plane of horse met-myoglobin (lattice B), taken with the Buerger precession camera. Angle of precession 17° , specimen-plate distance 8 cm., exposure time 105 hr., Cu $K\alpha$ radiation.

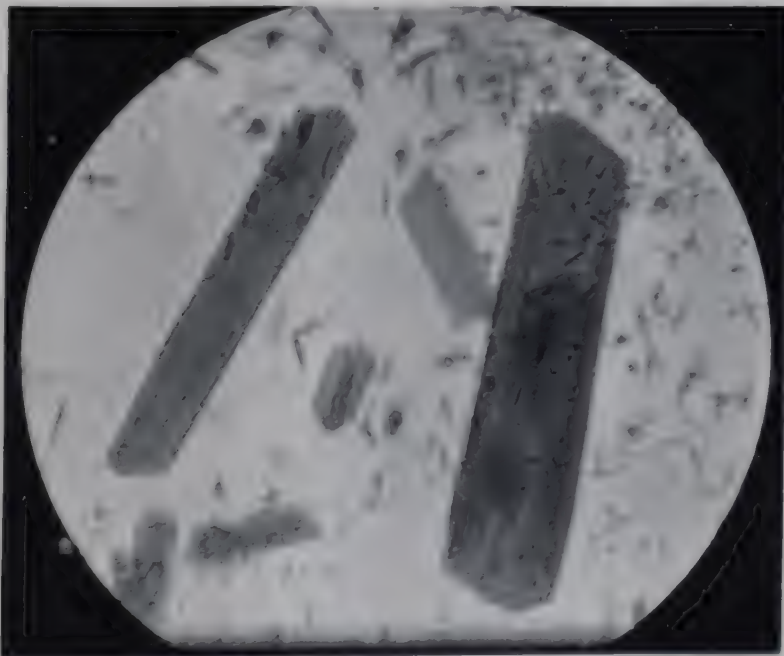
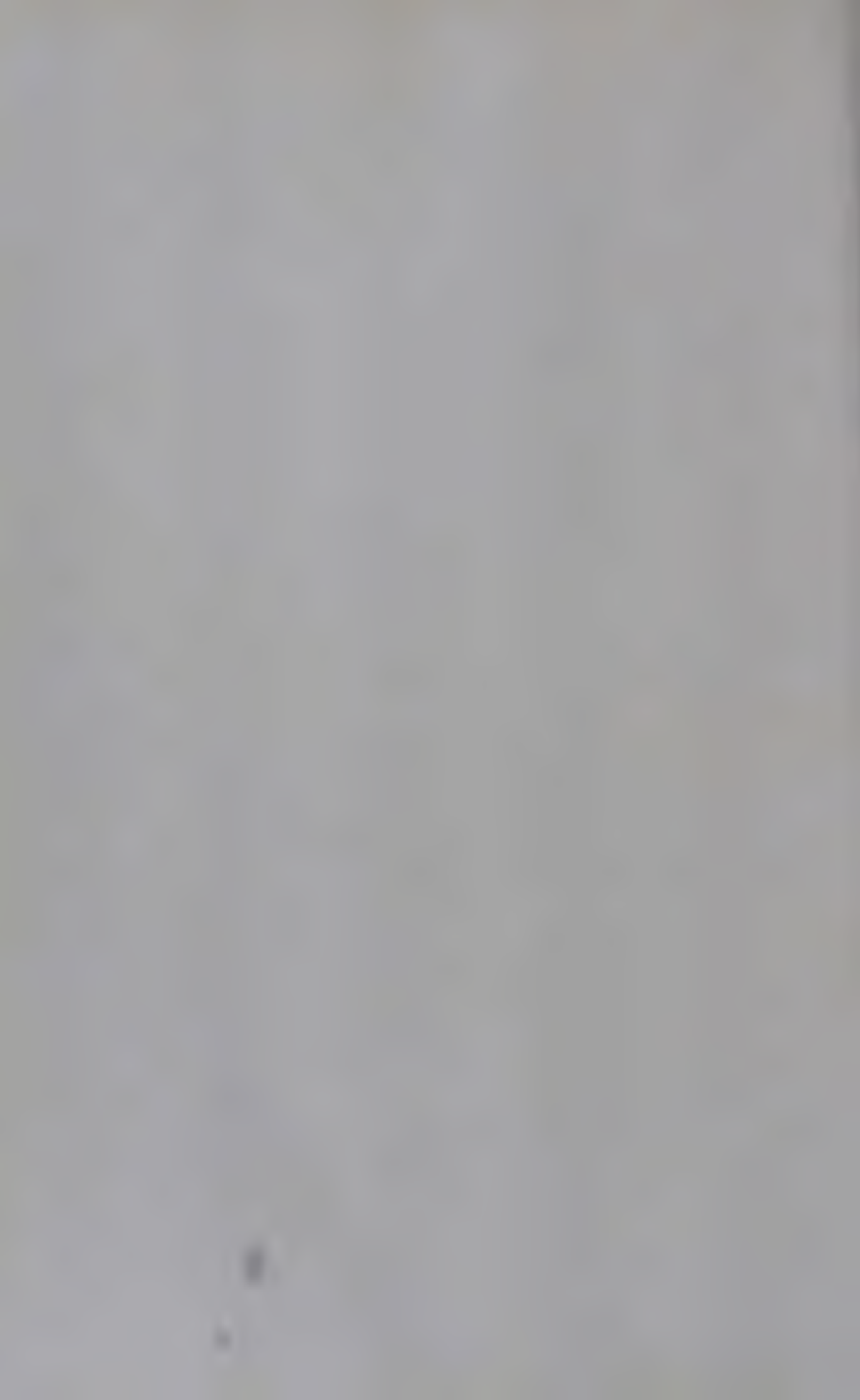


FIGURE 2. Crystals of horse met-myoglobin (magn. $\times 15$ approx.). (By courtesy of M. W. Rees.)



3. MORPHOLOGICAL AND X-RAY DATA

Figure 2 (plate 1) is a photograph of horse met-myoglobin crystals. They form long needles or lath-like plates, and they are monoclinic, the needle axis being b . The flat face of the lath is $\{001\}$, and the crystal is terminated by dome faces of two kinds, $\{110\}$ and $\{120\}$.

The optical properties of the crystals may be summarized as follows. Viewed perpendicular to the plate face there is straight extinction, the slow ray being parallel to b . Viewed in light parallel to b (i.e. with the crystal end-on to the incident beam) extinction is oblique, one extinction direction approximately bisecting β . By examining the birefringence of the crystal in various other orientations it can be shown that the vibration direction parallel to b is that of n_β ; the direction of n_γ is approximately the internal bisector of the monoclinic angle β . The crystals are optically negative; $2V = 30^\circ$. Pleochroism is strong, with high absorption parallel to n_β and n_γ , and weak absorption parallel to n_α .

All these relations are illustrated in figure 3.

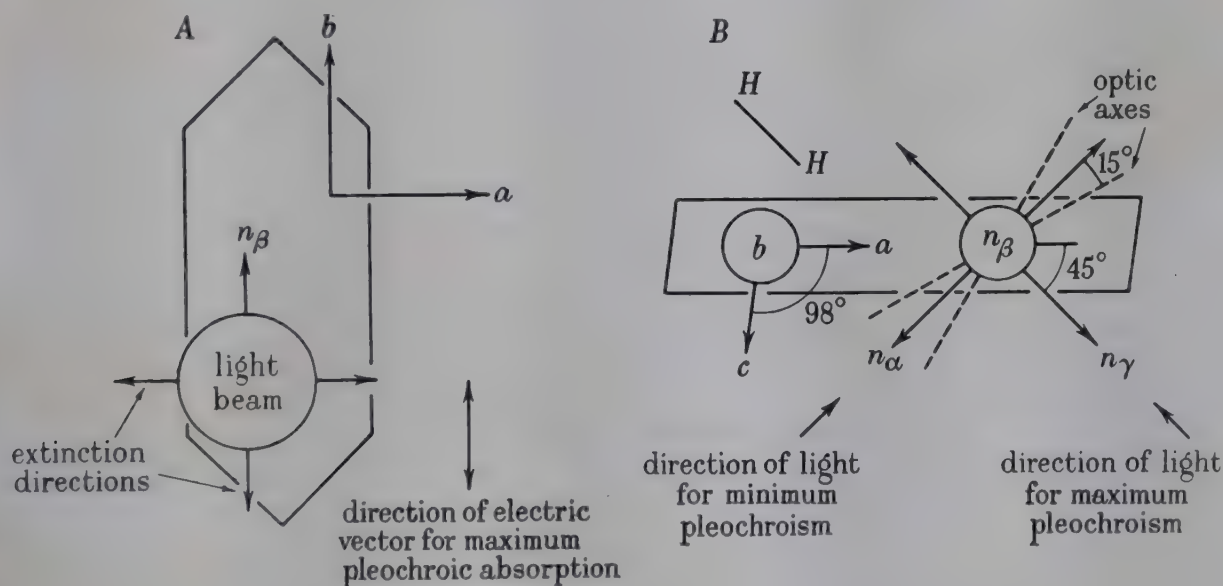


FIGURE 3. Optical properties of horse met-myoglobin (lattice B).
(A) Section parallel to ab . (B) Section perpendicular to b .

Cell dimensions and space group

The monoclinic unit cell was found to have dimensions $a = 57.3 \text{ \AA}$, $b = 30.8 \text{ \AA}$, $c = 57.0 \text{ \AA}$ and $\beta = 112^\circ$. The only systematically absent reflexion is $0k0$ when k is odd. Hence the space group must be $P2_1$ (if non-enantiomorphous space groups are excluded); it contains two general positions in the cell.

As in other crystalline proteins the cell dimensions alter in solutions of different salt concentrations or on drying. Owing to the small number of suitable crystals available it has not yet been possible to study these changes in detail. So far three different lattices have been encountered—in 3.5 M-phosphate, in saturated phosphate, and air-dried; the first lattice is that described above, and the dimensions of all three are given in table 1.

visual comparison with a standard scale, and, after applying a Lorentz factor, were used for computing a Patterson projection (corrections for polarization were negligible and were not applied). The corrected relative intensities and the projections derived from them are given in tables 2, 3 and 4 and figure 4 (lattice *B*); and table 5 and figure 5 (lattice *A*).

TABLE 3. RELATIVE INTENSITIES OF $0kl$ REFLEXIONS FROM LATTICE *B*

<i>l</i>	<i>k</i>										
	0	1	2	3	4	5	6	7	8	9	10
0	.	.	45	.	10	.	2	.	.	.	2
1	3	1	6	10	5	.	.	9	.	5	.
2	.	3	7	.	6	.	.	5	2	2	.
3	.	.	11	4	.	.	6	.	.	.	.
4	13	11	19	6	.	2	2	.	.	.	.
5	25	.	.	.	4	.	.	.	.	.	.
6	.	.	5	.	.	.	.	.	.	.	.
7	4	.	.	4	.	7	.	.	.	.	.
8	4	.	.	2	.	5	.	.	.	.	.
9	.	.	9	.	.	5	.	.	.	.	.
10	7	7	.	2	.	5	.	.	.	.	.
11	.	2	7	2	2	.	.	.	.	.	.
12	2	5	.	2	.	.	.	.	.	.	.
13	4	2	.	.	2	.	.	.	.	.	.

A brief digression is appropriate at this point to mention an important characteristic of Patterson (or vector) summations of macromolecular structures. This characteristic consists in a resemblance between the vector structure and the real molecular structure from which it is derived. For example, a series of equidistant parallel straight-chain molecules will give rise to a rod of vector density through the origin, surrounded by other parallel rods of vector density the same distance apart as the chains in real space; it is easy to see that the interatomic vectors within any one chain give rise to the rod through the origin, while the vectors between atoms of neighbouring chains produce the surrounding rods. Similarly, any arrangement in real space which projects on to a plane as a series of parallel equidistant bars of projected electron density gives rise to a projected vector structure (on the same plane) consisting of parallel rods of vector density, one of which passes through the origin. A more detailed discussion of these relations has been given by Kendrew & Perutz (1949); their main consequence is that, in favourable circumstances, there may be a general resemblance between vector structure and real structure. These effects may, of course, be obscured if the relative orientations of the molecules in the cell are unsuitable.

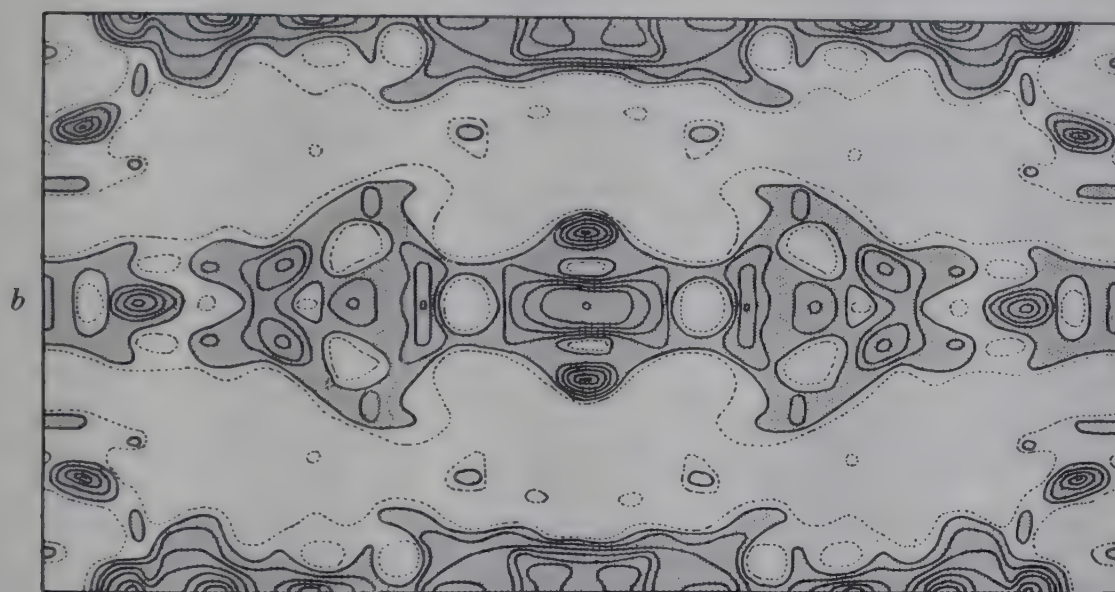
In both lattice *A* and lattice *B* the most prominent feature in the whole diffraction pattern is the very powerful 020 reflexion. This reflexion is mainly responsible for the striking appearance of the *c* and *a* projections, which will be described first.

(1) *c* and *a* projections, lattice *B* (figures 4 (a) and (b))

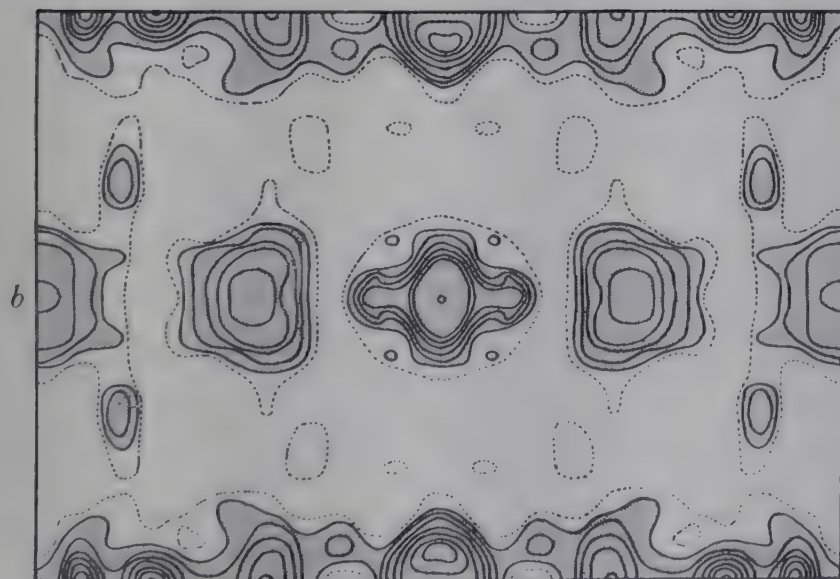
These projections are in many ways very similar. Each contains a very powerful ridge of vector density through the origin and perpendicular to *b*. Two further ridges run parallel to it at the top and bottom of the cell, and spaced from it by half the *b*

TABLE 5. RELATIVE INTENSITIES OF $h0l$ REFLEXIONS FROM LATTICE A

	← -h +h →																
	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8
0	.	.	.	.	2	3	7	4	.	4	7	3	2	.	.	.	.
1	.	8	.	7	.	.	2	4	4	.	9	.	2	7	10	.	.
2	.	8	15	16	13	4	5	.	.	1	4	.	.	.	.	.	.
3	11	7	9	.	5	3	.	2	.	3	3	.	.	.	.	.	.
4	3	8	.	4	4	.	1	1	.	14	32	.	.	.	9	.	.
5	3	5	10	.	10	5	5	.	8	.	12	8	.	.	17	.	.
6	.	6	.	.	20	.	2	9	29	10	3	.	.	.	.	.	.
7	.	6	.	.	.	.	10	.	.	.	.	.	.	.	.	.	.



(a) $a \sin \beta$



(b) $c \sin \beta$

0 Å 10

FIGURE 4(a), (b). Patterson projections of horse met-myoglobin (lattice B). Origins at centre of projections. (a) c projection. (b) a projection.

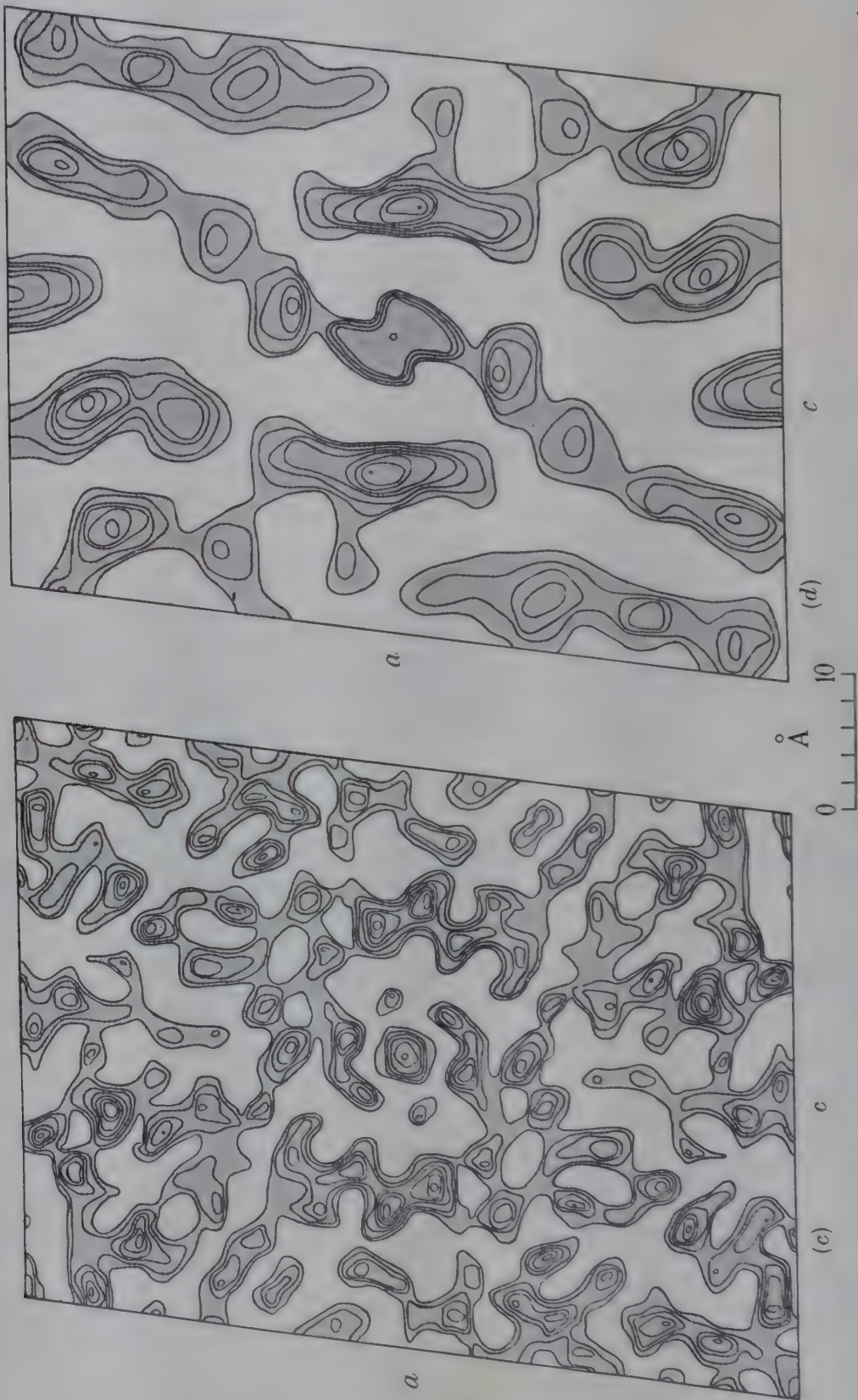


FIGURE 4 (c), (d). (c) *b* projection. (d) *b* projection, omitting from the summations all reflexions of spacing less than 7 Å.

dimension, i.e. $15.4 \text{ \AA}$. By virtue of the principle enunciated above it may be deduced that in real space the corresponding projections of the unit cell would exhibit strongly marked layers of high electron density, also perpendicular to b and $15.4 \text{ \AA}$ apart. The low vector density in the regions between the ridges shows that these layers of high projected electron density must be definitely separated by regions of lower electron density. Since both projections exhibit this layered structure, it must be concluded that the contents of the cell include two parallel layers, perpendicular

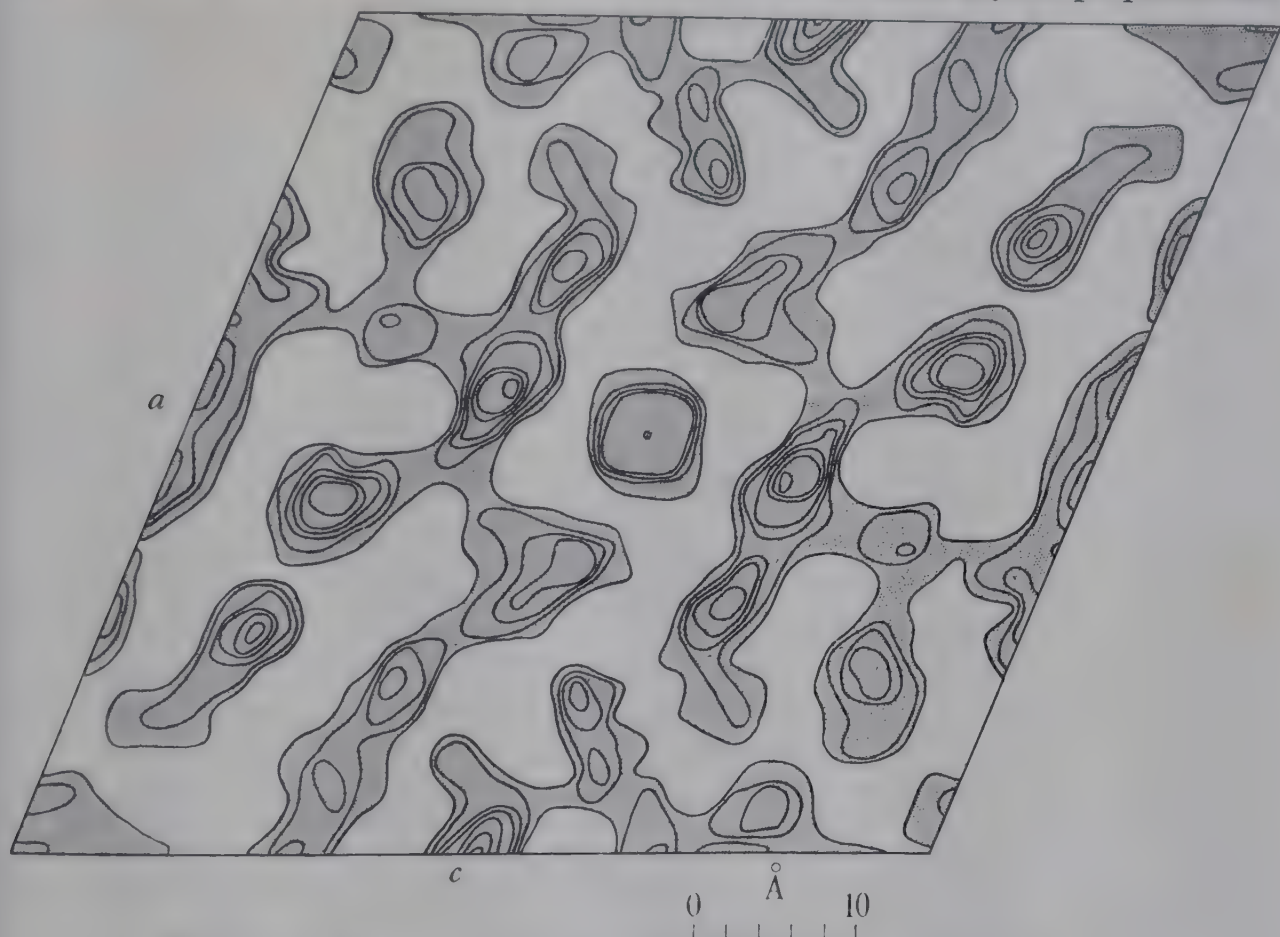


FIGURE 5. b Patterson projection of horse met-myoglobin (lattice A).
Origin at centre of projection.

to b and separated by $b/2$ or $15.4 \text{ \AA}$. The most plausible, though not the only possible,* interpretation of this state of affairs is to identify the two layers with the two myoglobin molecules which the cell contains, and to suppose them separated by a layer of liquid of crystallization (whose electron density will in general be different from that of the protein).

The a projection exhibits a further feature, namely, that each ridge of high vector density is segmented into prominent peaks, about $9.5 \text{ \AA}$ apart; there is no corresponding segmentation in the c projection. This must be due to an arrangement of the atoms in the layers (or molecules) in real space such that, viewed in a projection (but not in c projection), there are concentrations of high electron density about $9.5 \text{ \AA}$ apart.

A noteworthy feature of the c projection is the $4 \text{ \AA}$ vector parallel to b (see below, p. 78).

* See below, p. 82.

(2) *b* projections, lattice *A* (figure 5) and lattice *B* (figures 4(c) and (d))

Turning now to the *b* projections, it is best first to consider the one derived from lattice *A* (figure 5). This time the vector projection is of an entirely different type. There is first of all an irregular but distinct ridge through the origin at an angle of about 20° to the *a* axis, and extending throughout the width of the cell. Parallel to this and about $9.5 \text{ \AA}$ from it are two further ridges, and traces of yet another pair can be seen about $9.5 \text{ \AA}$ beyond these.

The deduction may be made that the real cell would yield a *b* projection consisting of parallel rods of high electron density, $9.5 \text{ \AA}$ apart. The existence of a screw dyad parallel to *b* shows that this projected structure must be due to two identical sets of rods in the cell, spaced $\frac{1}{2}b$ apart; presumably these sets of rods may be identified with the layers whose existence was deduced from the *a* and *c* projections. The rods in the lower layer will be parallel to those in the upper but in opposite orientation (since the screw dyad involves a rotation of 180°).

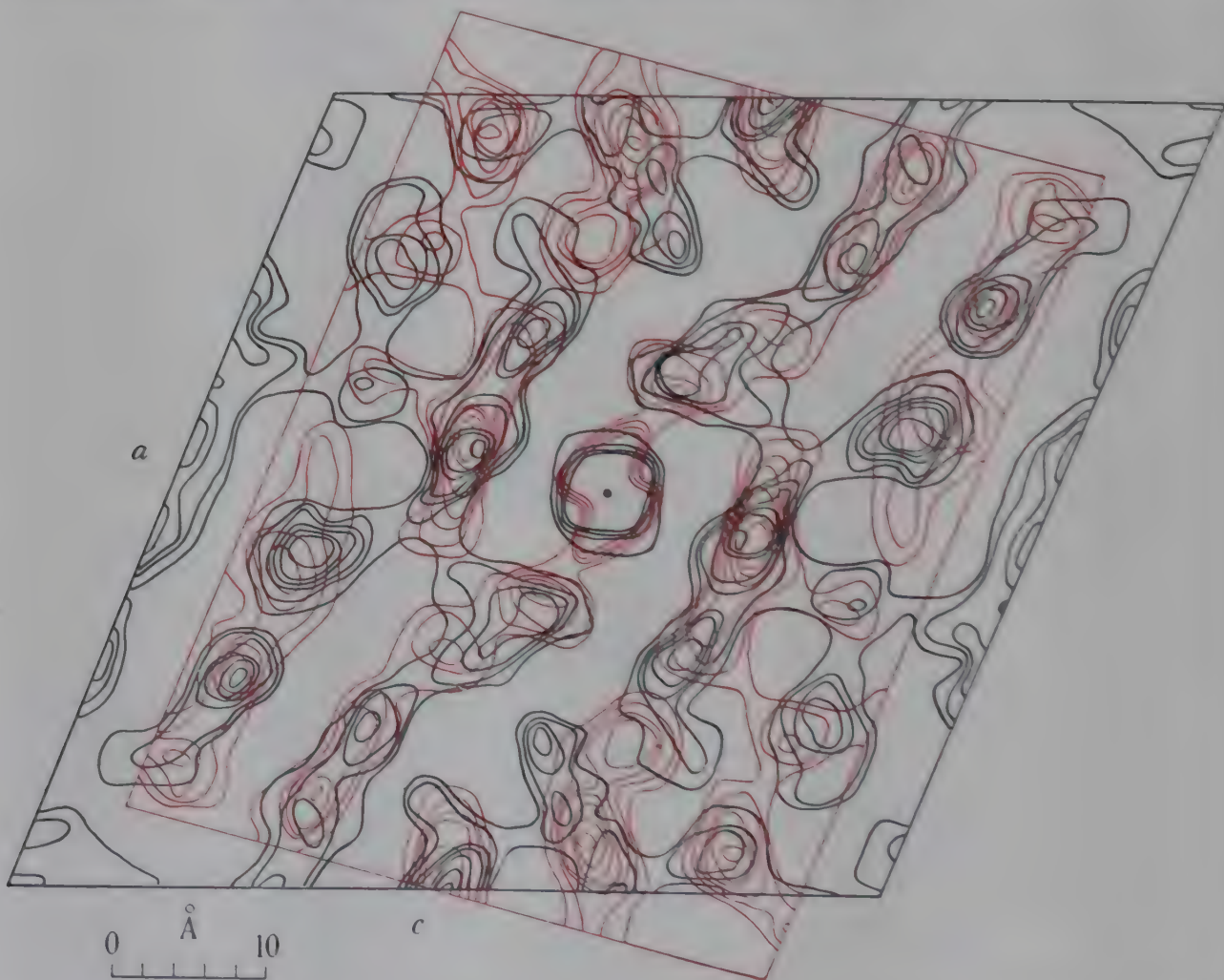


FIGURE 6. Superimposed *b* Patterson projections of horse met-myoglobin, lattices *A* and *B*.

In the *b* projection of lattice *B* (figure 4(c)) the vector structure is much more confused, far more peaks being visible and rods no longer being clearly distinguishable. It happened that the crystal from which the projection of lattice *B* was derived was larger and more perfect than that which yielded the lattice *A* pro-

jection, and gave X-ray photographs showing reflexions down to smaller spacings (about 2.5 Å). In order to obtain a projection more directly comparable to that of lattice *A*, and in order to reveal the more large-scale features of the vector structure, a second Patterson synthesis was computed, from the same X-ray photograph but including only the intensities of reflexions corresponding to spacings greater than 7 Å. This projection is shown in figure 4(*d*); it contains a very pronounced rod of high vector density through the origin, inclined at the same angle (20°) to *a* as the rods in the lattice *A* projection (figure 5). The rods displaced from the origin have, however, largely disappeared.

If figures 5 and 4(*d*) are superimposed (figure 6), it can easily be seen that many principal features are identical in the two projections. Moreover, certain peaks in figure 4(*d*) may be identified as remnants of the side rods in figure 5; and these fragmentary rods are separated from the origin rod by the same distance (9.5 Å) as those in figure 5. The suggestion may perhaps be made that the partial disappearance of the side rods in figure 4(*d*) is due to an increase in the number of *projected* rods of electron density on shrinking to lattice *B*. It is easy to see how this situation might arise by supposing that in lattice *A* the rods in the upper platelet lie almost, or quite, directly over those in the lower—if each platelet contains *n* rods the *b* projection in real space also contains *n* rods. During shrinkage to lattice *B* it may be supposed that the upper rods are slightly displaced relative to the lower, so that they no longer lie directly over them, and the *b* projection in real space contains 2*n* rods. It can be shown that in the general case the number of projected vector rods would then increase threefold in number, causing the vector projection to be more confused. Some evidence that the upper rods are not directly above the lower in lattice *B* is given below (p. 75).

Reverting to the *b* projection of lattice *B* derived from the complete diffraction pattern (figure 4(*c*)), attention is drawn to the set of very prominent 5 Å vectors arranged round the origin in a regular hexagon, and to the series of vector peaks 5 Å apart along each of the rods. (Neither feature is apparent in the lattice *A* projection which was derived from an X-ray photograph containing no reflexions of less than about 8 Å spacing.) The deduction to be drawn is that the rods in real space must have a repeating (or pseudo-repeating) structure of period 5 Å along them.

To summarize, the suggestion is that the two sets of parallel rods of high electron density in each cell are the two molecules of myoglobin; during shrinkage they alter their relative positions slightly, but the spacing of the rods within them remains the same. In other words, the sets of rods in the vector structure are the *intramolecular* patterns of the two molecules.

(3) *End-on Patterson projection of the rod structure*

The rods are inclined at an angle of only 20° to *a*, so a projection along *a* should show them nearly but not quite end-on. In fact the *a* projection discussed above (figure 4(*b*)) does exhibit a segmentation of the central layer, as has already been remarked; and this segmentation is not apparent in the *c* projection (figure 4(*a*)).

It will be noticed, in figure 4(*d*), that the rod direction is actually the same as that of the crystallographic direction $[20\bar{1}]$; that is to say, the rods lie parallel to the

(102) planes. This striking relation is true of both lattices *A* and *B*, and its significance will be discussed later (p. 80); in the meantime it will be apparent that it makes possible a Patterson projection which should reveal the rods *exactly* end-on.* Such an end-on view should be obtained by projecting the vector structure of the cell along the $[20\bar{1}]$ direction; the $[20\bar{1}]$ zone of reflexions can be recorded in the precession camera on a single photograph by making this direction the axis of precession.

TABLE 6. RELATIVE INTENSITIES OF $[20\bar{1}]$ ZONE OF REFLEXIONS FROM LATTICE *B*

		$h, \frac{1}{2}l$					
		0	1	2	3	4	5
<i>k</i>	0	.	1	29	3	4	4
	1	.	.	15	.	14	.
	2	32	7	.	.	7	4
	3	.	2	.	.	4	.
	4	5	.	.	4	.	4
	5	.	.	7	.	4	.
	6	2	.	.	.	.	.
	7	.	.	6	.	.	.
	8	.	.	4	.	.	.

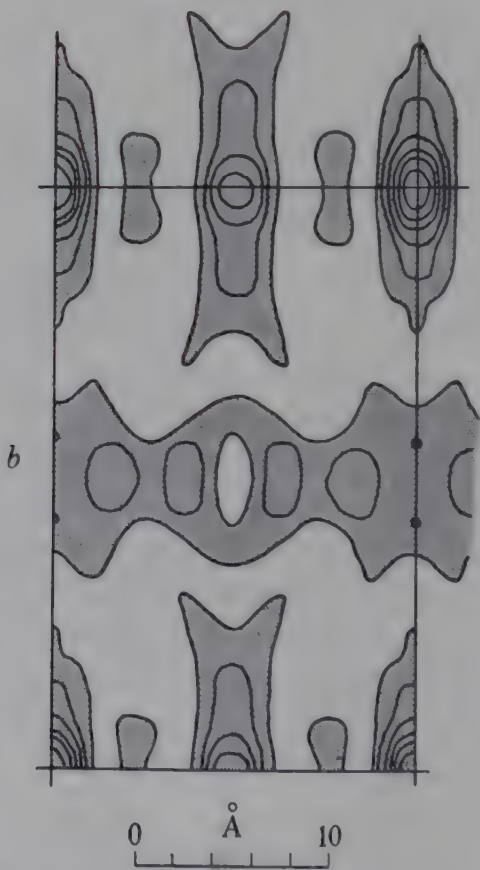


FIGURE 7. Patterson projection of lattice *B* of horse met-myoglobin along $[20\bar{1}]$. Origin at corner of cell.

The relative intensities of these reflexions (from lattice *B*) are given in table 6, and the Patterson projection derived from them is shown in figure 7. Its horizontal

* The author is indebted to Professor Sir Lawrence Bragg for making this suggestion.

repeat distance is, of course, the spacing between the (102) planes, namely $18.95 \text{ \AA}$; its other axis is b . As would be expected it exhibits the same general features as the a and c projections, namely, concentrations of vector density in layers perpendicular to b , and respectively passing through the origin and spaced $\frac{1}{2}b$ from it. The layer through the origin is built up of distinct peaks $9.5 \text{ \AA}$ apart, each repeating unit consisting of two of these peaks. However, the rods, now in end-on view, are very much better resolved than in the a projection; the height of the $9.5 \text{ \AA}$ vector peak is 58 % of that of the origin peak, as compared with 20 % in the a projection. This

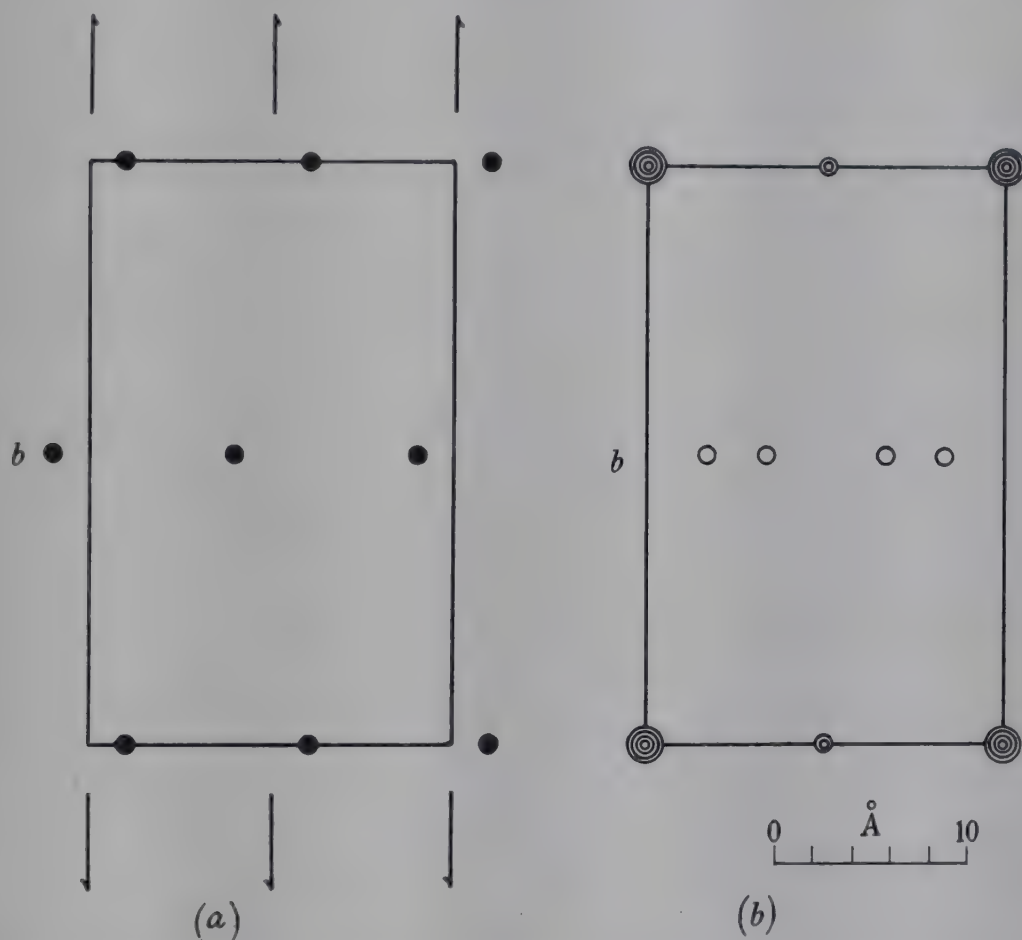


FIGURE 8. (a) Schematic arrangement of end-on projections of rods, viewed along $[20\bar{1}]$. (b) Patterson projection derived from (a). Origin at corner of cell.

arrangement, taken together with the evidence of the b projections, seems to settle beyond doubt the direction of the vector rods. Furthermore, the remarkably good definition of the peaks shows that in real space the rods from which they are derived must also 'stand out' very prominently from the rest of the structure; their mean electron density must be distinctly higher than that of the contents of the cell as a whole.

Turning now to the layer spaced $\frac{1}{2}b$ from the origin, it will be seen that it contains *four* peaks per repeat distance, instead of the two contained in the origin layer. Also the heights of these peaks are each about half those of the peaks in the origin layer. These relations are consistent with a staggered arrangement of the rods in real space, each rod being slightly displaced from a dyad axis. This is illustrated

schematically in figure 8(a), in which the projected rods are represented by circles. Figure 8(b) shows the Patterson projection derived from this arrangement; its resemblance to the Patterson projection of the actual crystal is obvious.

5. THE DIFFRACTION PATTERN OF THE DRY CRYSTAL

Like other protein crystals, those of myoglobin shrink on drying in air at room temperature, and the dried crystals yield relatively imperfect diffraction patterns. Owing to the poor quality of the crystals available it has so far been impossible to study these dry crystal diffraction patterns in detail, and they have not yet been fully interpreted. Photographs taken with the X-ray beam along b contained very few $h0l$ reflexions, and those very confused; in the other two axial directions, however, well-defined diffraction patterns were obtained, containing reflexions of spacings down to about 8 Å.

The following are the salient features of the dry diffraction pattern:

(1) 020 is a relatively weak reflexion, unlike the very strong 020 reflexion of the wet crystal. In other words, the pronounced layering of the cell perpendicular to b has now disappeared.

(2) The most striking feature of the whole pattern is the 001 reflexion, which is overwhelmingly strong. The wet crystal does not give a strong 001, and it must be concluded that on drying the molecules have regrouped themselves in such a way that they form sheets, c or 37 Å apart, parallel to the ab plane.

(3) The value of b in the dry crystal is 28 Å, that is, 2.8 Å less than in the wet crystal. The small degree of shrinkage is perhaps surprising if, as surmised in the last section, the wet crystal structure is built up of layers of molecules perpendicular to b and separated by liquid of crystallization. However, the wet crystals so far give no clue to the actual thickness of the molecules; and the observation in (1) above suggests that on drying something more complicated happens than a mere closing of the gap between the molecules.

(4) As already mentioned, the ac plane of the reciprocal lattice contains only a few reflexions, all of spacing greater than 25 Å. Some features of this plane are significant. The first few orders of $00l$ are well defined (including the very strong 001). The only other reflexions visible are $10l$'s and $20l$'s, but these rows are very confused and separate spots cannot be distinguished; it is as if the distribution of intensity along them were continuous.

(5) Perhaps connected with the above observation is the fact that it has so far proved impossible to find out whether β (chosen to correspond, in its relation to the protein molecule, with the value taken in the wet crystal) is obtuse or acute in the dried crystal. The diffracted patterns are consistent with a value of *either* 100° or 80°, and it may be that some kind of statistical twinning is involved. The matter requires further investigation when better crystals are available.

It was not practicable to prepare a b Patterson projection (see (4) above). Of the other two projections (not reproduced here), that along c contains no significant features, while the a projection exhibits a very powerful rod of vector density through the origin and parallel to b —the consequence of the very strong 001 re-

flexion. It also shows some traces of $9.5\text{ \AA}$ vectors perpendicular to b , perhaps corresponding to the similar vectors in the a projection of the wet crystal.

6. DISCUSSION

In studying the structure of myoglobin, or for that matter of any other crystalline protein, it is necessary to settle two important questions at an early stage:

(1) What is the arrangement of the polypeptide chains in the structure; that is to say, how are they oriented relative to the crystal axes and situated relative to one another, and what is the spatial configuration of the atoms making up the chains themselves?

(2) What is the shape of the molecule, and how are the different molecules in the cell related to one another?

These questions will now be discussed in turn. It may be objected that it is putting the cart before the horse to consider them in the above order; that it would be better to establish the general shape of the molecules and their mode of packing into the unit cell before beginning to consider features of their *internal* structure. The reverse order has been adopted because, of the evidence so far available, that which relates to the arrangement of the polypeptide chains is much less inconclusive than that relating to the shape of the molecule. It is shown below (p. 78) that the mean electron density inside the polypeptide chains is very much greater than elsewhere in the structure; the part of the molecule which consists of side-groups has an electron density not far different from that of the liquid of crystallization. Hence it is much easier to discriminate between the polypeptide chains and the rest of the cell contents, than between side-groups and liquid—and this latter type of discrimination is a necessary preliminary to settling the shape of the molecule.

The polypeptide chains

Neither time nor the quality of the crystals available have so far permitted the collection of the data which would be needed to prepare a three-dimensional Patterson synthesis of the myoglobin crystal. However, the Patterson projections along the three crystal axes and along the $[20\bar{1}]$ direction suggest strongly that the most prominent features of the three-dimensional structure in vector space would be two sets of rods of high density oriented along $[20\bar{1}]$. These sets, being related by the screw dyad parallel to b , would be $15.4\text{ \AA}$ apart, and each would consist of a number of parallel coplanar rods separated by $9.5\text{ \AA}$ and containing prominent peaks of vector density about $5\text{ \AA}$ apart along the rod direction.

The relations between the real and vector structures of macromolecular crystals suggest that the structure of the crystal in real space could be described in almost identical terms, substituting 'electron density' for 'vector density'. The only rod-like structures whose presence there is reason to suspect in a protein molecule are the polypeptide chains themselves, and it seems impossible to avoid the conclusion that the prominent features of the Patterson projections are in fact the projected vector structures of the polypeptide chains, whose orientation and arrangement are as set down in the last paragraph.

The striking fact about the $[20\bar{1}]$ projection, which may now be regarded as an end-on view of the chains, is the height of the peaks against a relatively uniform background. This implies that in real space an end-on view of the chains would show peaks of high electron density in the chains themselves, embedded in a fairly uniform background of lower electron density. This picture accords well with general notions about the structure of proteins. As much as half the molecular weight of a typical protein molecule is contributed by the chains themselves, the other half being made up by the side-groups. Furthermore, within the chains it would be expected that all (or nearly all) the atoms would be bound by covalent bonds; on the other hand, between the chains there would be regions containing large numbers of side-groups, molecules of liquid, etc., all built up of covalent bonds like the chain itself, but each being separated from its neighbour by Van der Waals bonds instead of covalent ones. The Van der Waals radii of atoms are between 2 and 3 times their covalent radii; the cube of this ratio (between 8 and 27) would be the ratio of the electron density of an assembly of atoms all separated by Van der Waals bonds, to that of an assembly all separated by covalent bonds. These figures set an upper limit to the ratio to be expected in the present instance, since many covalent bonds are also present in the inter-chain regions. The fact that the vector rods do appear as such prominent features of the Patterson projections leads to the important conclusion that not only must the polypeptide chains be parallel to one another, but their configurations must be such that under low resolution they appear as more or less straight rods—a meandering chain would not be detected in spite of its high electron density. In the arrangement shown in figure 8(a) Bragg (1949) has calculated that for one chain model satisfying these criteria (actually a trigonal spiral) the projected electron density within the chains would be about 10 times that outside them.

Another significant point is the shape of the peaks in the $[20\bar{1}]$ projection (figure 7). It will be seen that the origin peak and its companion at the same height are considerably drawn out in the b direction; in fact there are signs that each is made up of three imperfectly resolved peaks, one main central one and two satellites each spaced about $4\text{\AA}$ away from it (similar satellite peaks are visible below and above the origin in the c projection, figure 4(a)). Such an arrangement would be produced by an array of *pairs* of peaks of projected electron density in real space, the members of each pair being $4\text{\AA}$ apart. This scheme is illustrated in idealized form in figure 9(a); figure 9(b) is the Patterson projection derived from this model. Its consequences have been further explored by using it to calculate the phases of the reflexions in the $[20\bar{1}]$ zone; these calculated phases, together with the experimentally observed intensities, were then used to calculate a *Fourier* projection which is illustrated in figure 10. Some of the observed reflexions were perforce omitted from this calculation: the model possesses symmetry elements—mirror and glide planes perpendicular to b and $[20\bar{1}]$ respectively—which are absent from the real structure, and therefore gives $F(hk) = 0$ for h odd, so no clue is obtained about the signs of the reflexions having h odd. However, there are not very many such reflexions, and those which do occur are weak (suggesting that the real structure has a tendency to a pseudo-glide plane, which is some confirmation of the model);

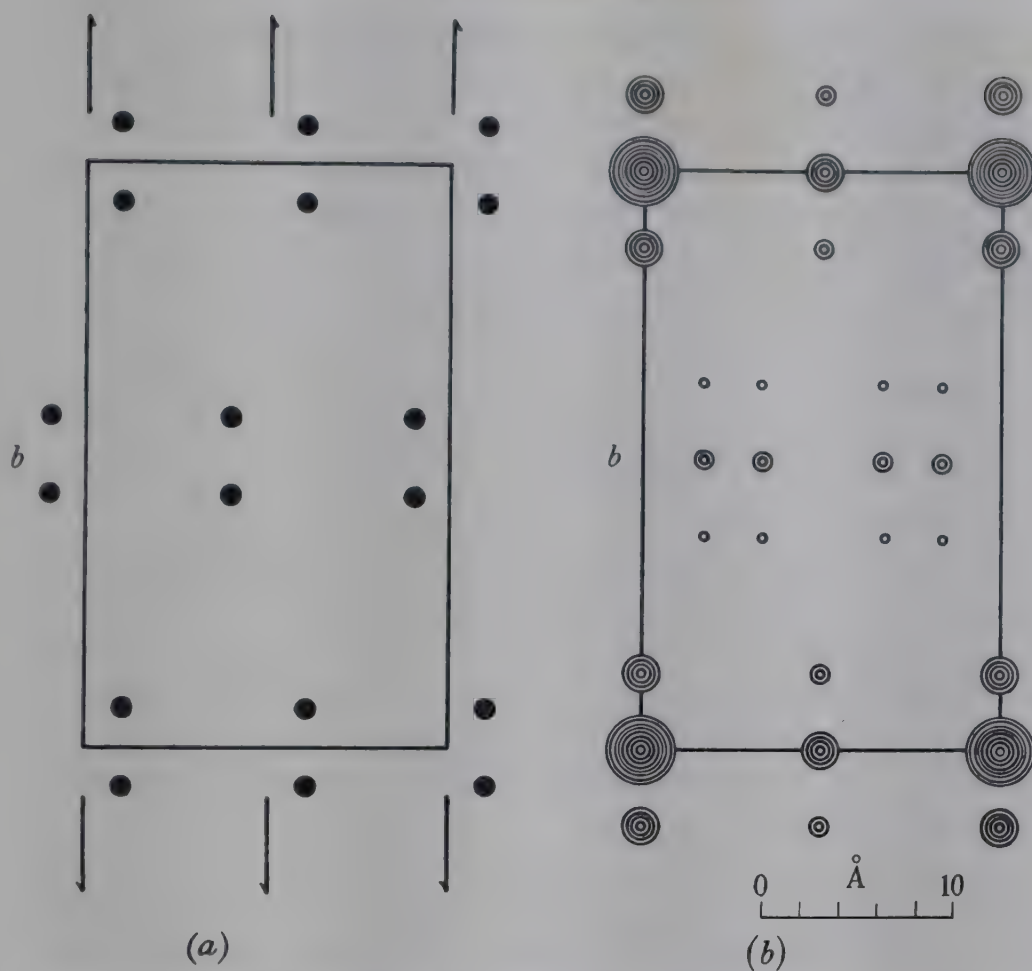


FIGURE 9. (a) Schematic arrangement of end-on projections of *double rods*, viewed along $[20\bar{1}]$. (b) Patterson projection derived from (a), for comparison with figure 7. Origin at corner of cell.

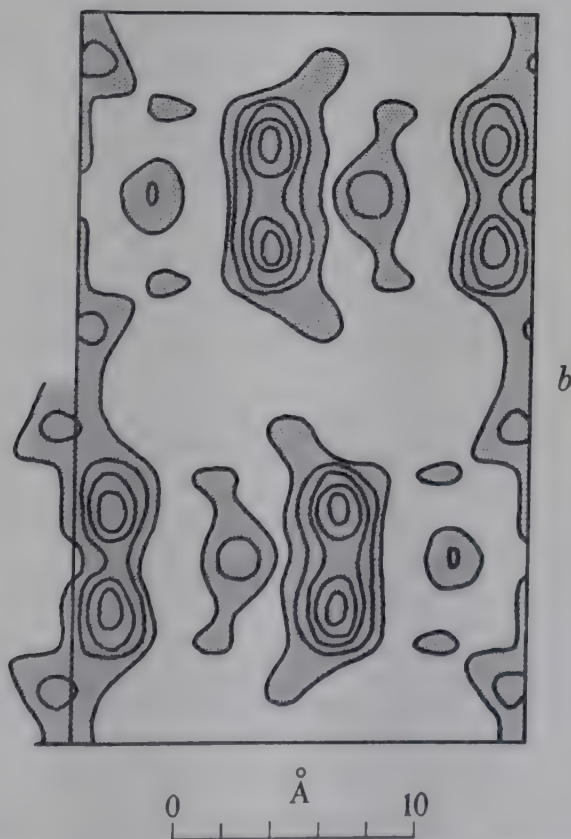


FIGURE 10. Fourier projection along $[20\bar{1}]$, using intensities determined experimentally and phases calculated from the model of figure 9 (a).

in fact, one such reflexion, with arbitrary sign, could be included—the strongest being chosen—while the rest were left out entirely. The omission of a few of the weakest reflexions probably has little material effect on the result.

It must be admitted that the process by which figure 10 was arrived at is somewhat arbitrary; and in strict logic all that can be claimed is that it is merely one of many structures compatible with the observed diffraction pattern. However, general notions about the relations between real and vector structures suggest that the most striking features of this Fourier projection must be common to *all* such compatible structures. It should be remembered that, like the model from which it was derived, it has a higher symmetry than the real structure, possessing mirror planes perpendicular to b , and, apart from the perturbations introduced by the solitary reflexion with odd h , glide planes perpendicular to $[20\bar{1}]$.

The considerations mentioned in the last two paragraphs suggest that the end-on projection of the polypeptide chains is not even approximately centro-symmetrical, but that it is drawn out in the b direction and perhaps definitely split into a pair of concentrations of projected electron density each of which, however, might be approximately centro-symmetrical. Remembering that we are dealing throughout with projections of the cell, and not with sections through it, this appearance might have any of the following causes:

(1) That each chain is thrown into a series of short-range folds (in a plane defined by b and $[20\bar{1}]$) seen end-on in this projection; this kind of structure has been postulated by Astbury & Bell (1941) for α -keratin. This is to say that the unsymmetrical projection of the chains is due to their unsymmetrical cross-section.

(2) That any *section* through the chain is centro-symmetrical, but *either* (i) the projection is elongated because the chains do not precisely follow $[20\bar{1}]$ through their whole length, being displaced from it by small amounts in the b direction at various points; in other words, the chains are slightly serpentine, *or* (ii) the $[20\bar{1}]$ projection derives from different molecules, or regions of a molecule, in which sections of chain are alternately displaced *up* and displaced *down* from the mean position, though each section is straight and accurately follows the $[20\bar{1}]$ direction.*

A choice between these alternatives would provide valuable evidence of the detailed configuration of the polypeptide chains; unfortunately, it is not possible to discriminate between them on the basis of the two-dimensional evidence now available.

The shrinkage of the unit cell

Examination of the shrinkage of the crystal revealed the following striking relations:

(1) In both lattice A and lattice B the chain direction is the same as the crystallographic direction $[20\bar{1}]$.

(2) On shrinkage the angle between a and $[20\bar{1}]$ remains almost constant, and the distance between successive (102) planes (which is double the chain separation) changes very little.

* In the end-on projection one is 'looking' at a total depth of $128\text{ \AA}$; the projection is two unit cells deep.

These facts suggested that it might be profitable to outline the unit cell in a different way, as shown by the dotted lines in figure 11. This procedure is of course permissible, since a monoclinic cell is not characterized by a unique pair of a and c axes; the dotted cell, like the original one, is primitive and has screw dyad axes at the corners and centre of its b face. The new axes may be called a' and c' , where c' is the old $[20\bar{1}]$ direction and a' is the same as the old a . The distance between the old (102) planes becomes $d'_{(100)}$. The old and the new cells may be designated (abc) and $(abc)'$ respectively.

It is easy to calculate the dimensions of the $(abc)'$ unit cell for each shrinkage stage; this has been done in table 7, where θ is the supplement of β' (the monoclinic angle of the new cell) and is also the angle between the old $[20\bar{1}]$ direction and a .

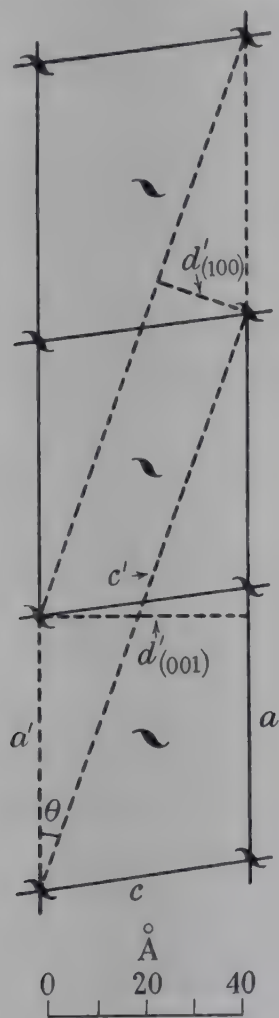


FIGURE 11

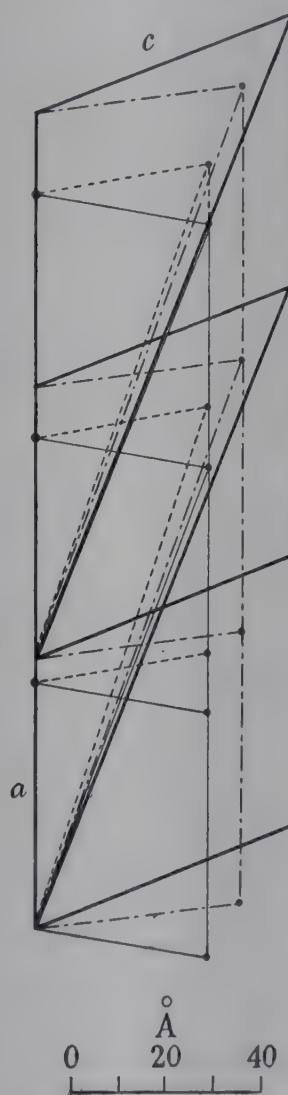


FIGURE 12

FIGURE 11. Alternative method of outlining the b face of the unit cell. Original axes a , c ; new axes a' , c' .

FIGURE 12. Shrinkage of the b face of the unit cell, showing that the (102) planes maintain their direction and separation approximately constant. —, lattice A; - - -, lattice B; - - - -, dry lattice, taking $\beta = 100^\circ$; —, dry lattice, taking $\beta = 80^\circ$.

The table includes two sets of values of the dimensions of the dry $(abc)'$ cell, worked out on the basis of $\beta = 100^\circ$ and $\beta = 80^\circ$ respectively; the differences between

the two sets are not much greater than the experimental errors in the determination of dry cell dimensions.

TABLE 7. DIMENSIONS OF SHRINKAGE STAGES OF $(abc)'$ CELL

	a'	b'	c'	θ	$d'_{(100)}$	$d'_{(001)}$
lattice A	57.3 Å	30.8 Å	146 Å	21.2°	20.7 Å	52.8 Å
lattice B	57.3	30.8	128	19.5°	19.1	42.7
dry (β obtuse)	(51.5)	(28.0)	(114)	(18.7°)	(16.5)	(36.6)
dry (β acute)	(51.5)	(28.0)	(103)	(20.8°)	(18.3)	(36.6)

It will be seen that the $(abc)'$ cell shrinks almost entirely by means of a decrease in the value of c' ; the transition from A to B leaves a' and b' constant, and θ almost so. These relations are illustrated in figure 12; and they suggest strongly that

(1) The major part of the liquid of crystallization is situated on the c' faces of the molecules, and very little on the a' or b' faces.

(2) The molecule consists of lengths of chain parallel to c' and has more or less flat sides parallel to this direction. Only thus can it be understood how, when the crystal shrinks, one molecule can slide past its neighbours without appreciably altering any dimensions of the lattice except c' .

(3) Even in the most expanded lattice the molecules are more or less in contact along the a' direction, since in shrinking from lattice A to lattice B the value of a' remains quite unchanged.

The shape of the molecule

The data so far obtained are not sufficient to determine the general outlines of the molecule unequivocally. At present all that can be done is to consider what types of molecule are compatible with the geometrical relations just discussed, and with the Patterson projections.

It will be remembered that every Patterson projection made along a direction perpendicular to b contains horizontal regions of high vector density through the origin and at a distance $\frac{1}{2}b$ away from it. This strongly suggests that the contents of the cell are arranged in layers of alternately high and low electron density, $\frac{1}{2}b$ apart and perpendicular to b ; and the implication would be that the two molecules in the cell are flat platelets, related by a screw dyad and with their flat surfaces perpendicular to b . If this is so then each molecule must be built up of a set of parallel polypeptide chains all more or less in the same plane. This is the most reasonable interpretation of the Patterson projections; it is supported by the evidence that the relative positions of the sets of chains in the two layers change during shrinkage, both from side to side and end to end. However, the possibility cannot absolutely be excluded that the molecule consists of two layers of chains instead of one; the calculation of electron density within and outside the chains shows that it is the position of the latter, rather than the boundary between molecule and liquid, which determines the main features of the Patterson projections.

The next point to consider is what types of outline the molecule might possess when viewed along b ; in other words, what shapes in the ac plane are compatible with the features of the b Patterson projections and with the deductions made, at the

end of the last section, from the geometrical relations between the various shrinkage stages. One such shape is illustrated in figure 13(a), which shows the molecules in lattice *B*. This model is of a rectangular molecule built of four lengths of polypeptide chain; each cell would contain a second molecule underneath the first and related to it by a screw dyad. Thus the molecules would have the 'platelet' form discussed in the last paragraph; they would be only one chain deep. Expansion and contraction of the lattice would take place by a sliding of the molecules past one another along $[20\bar{1}]$; in lattice *A* the molecules would be farther apart along the chain direction, as shown in figure 13(b), while in the dry lattice they would approach more closely together than in lattice *B*. It will be noted that the molecules are

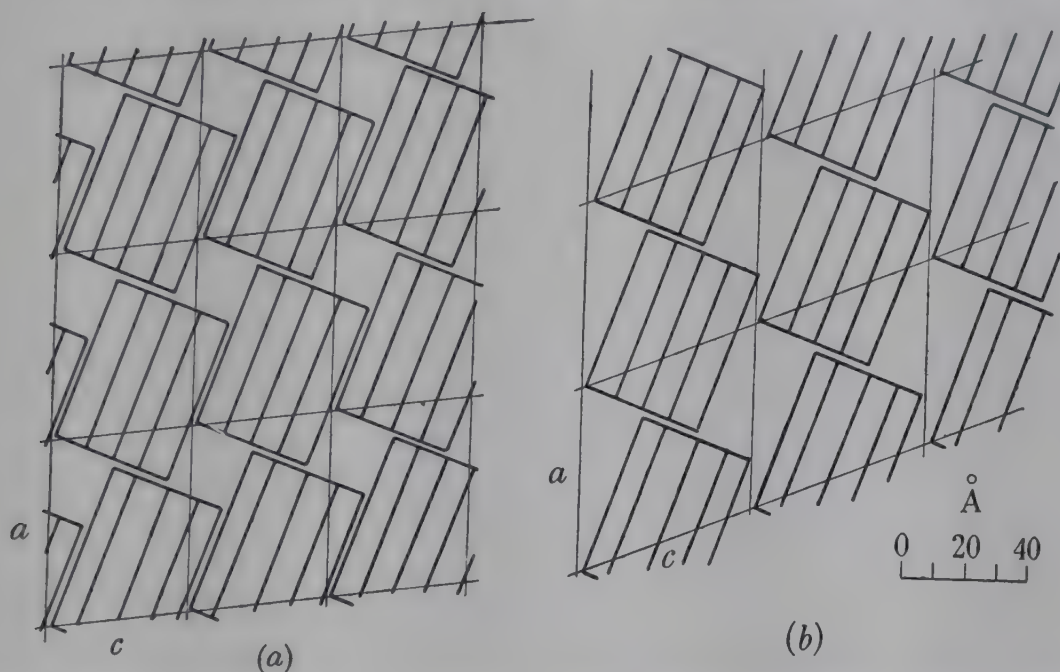


FIGURE 13. Possible model of myoglobin molecule, built up of four parallel lengths of polypeptide chain. Viewed along *b*. (a) In lattice *B*. (b) In lattice *A*.

grouped in rows parallel to a' in such a way that on shrinkage each row slides past its neighbours without its own structure being altered; this arrangement would account for the fact that a' has the same value in lattices *A* and *B*, though in the dry lattice it seems to be distinctly shorter, indicating that in the last stage of shrinkage some distortion of the structure has taken place. Possibly the rows are further consolidated in this last stage, forming a structure which would account for the very powerful 001 reflexion from the dried crystal. The strength of this reflexion, which implies that the molecules are in sheets perpendicular to *c* and 37 Å apart, can be reconciled with the idea of rows of molecules in contact along *a* in the wet state; on drying these rows presumably fall above one another and form coherent sheets of molecules in the *ab* plane. It has already been remarked that the *a* projection of the dry crystal shows traces of the interchain vectors of 9.5 Å.

This model fits in reasonably well with either value of the dry β (80° or 100°). In each case θ remains in the neighbourhood of 20° , and $d'_{(100)}$ has a value not far different from that in lattices *A* and *B*. On balance $\beta = 80^\circ$ would be easier to reconcile—the dimensions of the dry lattice then correspond exactly to the condition

where the gaps between the c' faces of the molecules in figure 13 have closed up altogether—but the uncertainties in the dry crystal data are such that no definite choice can be made.

The model illustrated in figure 13, consisting of a flat pack of four parallel polypeptide chains (one chain deep), is merely one of a number compatible with the data. It is included to illustrate some features which the molecule must possess, and because it is perhaps more plausible than any other. Further data will be required before more definite conclusions can be reached. Meanwhile it may be noted that the platelet model would have an asymmetry of 3.5:1 or 4:1, agreeing well with the findings of Wyman & Ingalls discussed in §1, but not with those of Gutfreund. The latter author prefers a spherical, or almost spherical, shape; but although such a molecule could just be fitted into the unit cell it is hard to understand how an assembly of close-packed spheres could give Patterson projections of the type observed. In any case comparisons between crystal data and measurements of asymmetry in solution cannot be very profitable until both have been put on a firmer basis.

The orientation of the haem group

As in haemoglobin it is plausible to suppose that the absorption in the visible region of the spectrum, which confers on myoglobin its colour, is due to the conjugated ring system of the haem group, and that it reaches a maximum when the electric vector of the incident light lies in the plane of that ring system. The strong pleochroism of myoglobin crystals shows that all the haem groups in the structure must be more or less coplanar. Furthermore, the observation that pleochroism is a maximum when the incident light lies (approximately) in the direction of the internal bisector of β , and that in this orientation the absorption is a maximum when the electric vector coincides with b , shows that the plane of the haem groups is that containing b and the internal bisector of β ; this arrangement is illustrated in figure 3, where $H-H$ indicates the direction of the haem groups.

Turning now to the refractive indices, it has already been noted that the crystals are optically negative; the two higher, and nearly equal, refractive indices n_β and n_γ lie in the plane allotted to the haem groups, while the lowest index n_α is perpendicular to it. These observations, and the smallness of the difference between n_β and n_γ , indicate that the haem group exerts a dominant influence in determining the birefringence of the crystals.

It has been suggested that the molecules of myoglobin may be platelets lying perpendicular to b ; the haem groups, on the other hand, lie in a plane containing b . The haem group would therefore be oriented at right angles to the platelet. The direction and magnitude of the two higher refractive indices are in conformity with such an arrangement, for the higher of the two (n_γ), representing the greatest retardation, lies in the direction containing both the haem group and the plane of the protein platelet; the lesser (n_β) lies in the plane of the haem but perpendicular to that of the protein.

Reference to the ac projection (figure 5) shows that the haem group, which roughly bisects β , must also be more or less perpendicular to the direction of the polypeptide chains.

Relations with haemoglobin and α -keratin

The structural data for myoglobin display analogies to the model proposed by Perutz for horse haemoglobin (Boyes-Watson, Davidson & Perutz 1947; Perutz 1949). He has concluded that, broadly speaking, the haemoglobin molecule is a cylinder, 34 Å deep and 57 Å across; within it the polypeptide chains are arranged in four layers perpendicular to the axis of the cylinder, and each layer is just under 9 Å thick and consists of a set of coplanar polypeptide chains, parallel to one another and about 10 Å apart, and exhibiting repeat distances of 5 Å along the chain direction, each repeating unit consisting of three amino-acid residues. The four haem groups are parallel to one another and lie in a plane perpendicular to the polypeptide chains. Their positions in the molecule are unknown, though it is natural to suppose that one haem group is associated with each of the four layers.

The first hint of a close relationship between the structures of haemoglobin and myoglobin came from the remarkable correspondence between the postulated diameter of the haemoglobin molecule (57 Å) and the dimensions of the lattice *A* unit cell of myoglobin ($a = 57$ Å, $c = 57$ Å). In fact a circle of diameter 57 Å very nearly fits into the *ac* face of the unit cell, since β is not far removed from 90° (see figure 14). Another way of looking at this relation is to compare the dimensions of the dry crystal (in which neighbouring molecules are presumably in contact) parallel to the chain direction in each protein. In haemoglobin this is the *a* axis, which is 102 Å in the dried crystal; in myoglobin the corresponding axis is c' , which shrinks to about 103 or 114 Å, according to the value taken for β .

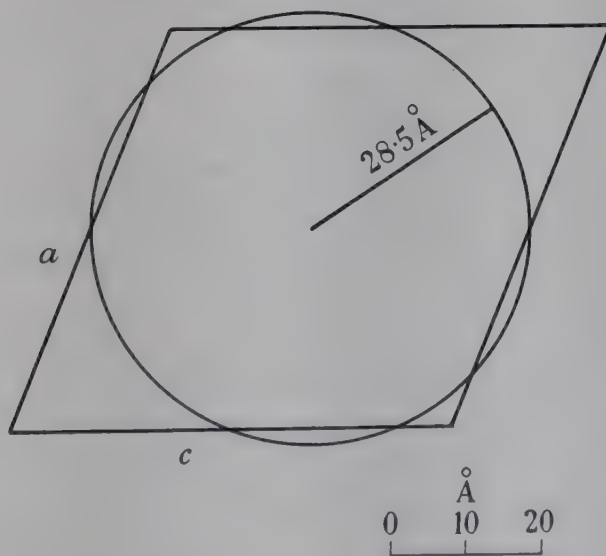


FIGURE 14. Circle of diameter 57 Å superimposed on the *ac* face of the unit cell of myoglobin (lattice *A*).

The resemblance goes further, for the vector structures of the two proteins display striking similarities. In each case there are parallel rods of vector density, about 10 Å apart and containing prominent vector peaks 5 Å apart along the rod direction. Furthermore, if the *b* projection of myoglobin (figure 4(c)) is superimposed on the zero-level *b* Patterson section of haemoglobin (Perutz 1949, figure 2)

so that the rod-like structures through the origin lie on top of one another, it can be seen that many of the peaks along these structures almost coincide.

The 'platelet' model of the myoglobin molecule consists of a flat pack of parallel and coplanar polypeptide chains about 10 Å apart, and associated therewith a haem group perpendicular to the chains, while along the chains is a series of 5 Å repeats. This arrangement would also describe one of the four layers in Perutz's model of horse haemoglobin.* When all these observations are taken together with the fact that, within the limits of experimental error, the molecular weight of myoglobin is one-quarter of that of haemoglobin, it is attractive to suggest that myoglobin has a structure analogous to a single one of the four layers of haemoglobin, as regards the arrangement of the polypeptide chains, the general shape and dimensions of the molecule, and the orientation of the haem group.

Perutz (1949) has noted that there is a close resemblance between the structure of the polypeptide chains in haemoglobin and the chain structure in α -keratin (and related fibres) which is responsible for its wide-angle X-ray diffraction pattern. In each case there are parallel chains 10 Å apart, exhibiting a 'repeat' of about 5 Å along the chain direction. The resemblance between the chain structure in myoglobin and that in haemoglobin has already been commented upon. The similarity between the chain structures in haemoglobin and myoglobin on the one hand, and in the α -form of the keratin-myosin-fibrinogen group of fibrous proteins on the other, suggests strongly that the α -keratin type of folded polypeptide chain may be a fundamental structural feature of wide occurrence among all types of protein. It is unfortunate that the stereochemical lay-out of the fold characteristic of α -keratin has still to be conclusively established; in the meantime it is possible only to recognize the spacings characteristic of it.

Chemical data

Mention has been made in §1 of the result of applying the 'end-group' method to myoglobin, namely, that, excluding possibilities of cyclic arrangements, the molecule consists of a single continuous polypeptide chain. To accommodate this chain within the model suggested above, it must be arranged in a series of loops, the limbs of which are straight but might of course be of different lengths. Now in α -keratin and haemoglobin each main-chain repeat unit of 5.1 Å is thought to consist of three amino-acid residues; assuming that the same arrangement holds good in myoglobin, and bearing in mind that the molecule contains about 146 amino-acid residues in all, the total length of chain must be about $5.1 \times 146/3$ or 248 Å. In the model illustrated in figure 13 there are four segments of chain each $a' \cos \theta$ or 54 Å long, giving a total chain length of 216 Å. The agreement is not exact; however, the discrepancy is fairly small (amounting to 8 Å per segment) and might be accounted for by supposing that the molecule is not precisely rectangular or that some chain length is taken up in 'going round the corners', which must happen since the molecule consists of a single chain.

* The model proposed for myoglobin is rectangular, whereas the haemoglobin model has a circular section. These are not incompatible since neither model could claim to be more than a rough approximation to the actual shape of the molecule.

7. CONCLUSIONS

(1) Horse met-myoglobin crystallizes in the monoclinic space group $P2_1$. The unit cell contains two molecules of mol.wt. 17,000 occupying general positions. Each molecule contains one haem group.

(2) The crystals are strongly pleochroic, which suggests that all the haem groups in the structure are approximately parallel; and it is deduced from the directions of maximum absorption that they lie in a plane containing b and the internal bisector of β .

(3) The most striking feature in the diffraction pattern of the wet crystal is a very strong 020 reflexion. The a and c Patterson projections accordingly contain a powerful rod of vector density passing through the origin and perpendicular to b ; two other rods run parallel to it at the top and bottom of the cell, $\frac{1}{2}b$ or $15.4 \text{ \AA}$ away.

(4) The b Patterson projection contains a well-marked rod of vector density passing through the origin and inclined at about 20° to a . Four other rods run across the projection in parallel directions, two being $9.5 \text{ \AA}$ from the central one, and the other two $19 \text{ \AA}$ from it. The rods contain maxima $5 \text{ \AA}$ apart. The direction of the rods coincides with the crystallographic $[20\bar{1}]$ direction, and a Patterson projection along this axis shows them end-on and well resolved. The features of this projection are compatible with an arrangement in real space of the kind shown in figure 10.

(5) The vector rods are considered to be derived from two sets of parallel polypeptide chains. These chains run parallel to $[20\bar{1}]$, and lie $9.5 \text{ \AA}$ apart; they are built up of repeating units of period $5 \text{ \AA}$. The two sets of chains are related by the screw dyad symmetry of the cell.

(6) The unit cell may be outlined in the way shown by the dotted lines in figure 11. The a' , b' , β' and $d'_{(100)}$ dimensions of this cell change very little on shrinkage, while c' shrinks considerably. This is interpreted as meaning that the molecules are built up of chains parallel to c' (or $[20\bar{1}]$) and have flat sides also parallel to this direction; the liquid of crystallization is mainly situated on their c' faces. During shrinkage the liquid is withdrawn and the lattice contracts in consequence of the molecules sliding past one another along their flat sides. The identity of a' in lattices A and B is thought to mean that the molecules are more or less in contact, and form rigid rows, along this direction.

(7) The electron density inside the chains is much higher than elsewhere in the structure, and the prominent features of the Patterson projections are due primarily to this difference. The difference between the electron density of the rest of the molecule and that of the liquid of crystallization is much less. Hence the Patterson projections give much more information about the arrangement of the polypeptide chains than about the overall shape of the molecule. Neither the projections, nor the geometrical relations summarized in (6), give conclusive evidence. But on balance they favour a 'platelet' molecule built up of four parallel co-planar lengths of chain, each about $54 \text{ \AA}$ long; the width of the molecule would be about $37 \text{ \AA}$ and its depth between 9 and $14 \text{ \AA}$. In the unit cell the platelets would be oriented perpendicular to b and would be related by a screw dyad. The haem groups would

be perpendicular to the platelets and also to the polypeptide chains. This model is illustrated diagrammatically in figure 13.

(8) There are striking resemblances between the Patterson projections of myoglobin and the Patterson sections of horse haemoglobin, and between the unit cell dimensions of the two proteins. The haemoglobin molecule is believed to consist of a cylinder 57 Å in diameter and 34 Å deep, made up of four layers; each layer is just under 9 Å thick and consists of parallel polypeptide chains about 10 Å apart, while the four haem groups are oriented perpendicular to the chains. Furthermore, the molecular weight of haemoglobin is four times that of myoglobin. The tentative suggestion is made that myoglobin might be analogous in structure to one of the four layers of haemoglobin, and, since the amino-acid compositions of the two proteins are quite different, the implication would be that the resemblance between them in properties and in physiological function has its origin in such an analogy.

(9) In both myoglobin and haemoglobin the polypeptide chains are 9 to 10 Å apart and exhibit repeat distances of 5 Å along their length; it is suggested that the polypeptide chains are folded in a similar way in these two proteins, and that the same chain configuration may also occur in the α -keratin group of fibrous proteins, which is characterized by almost identical spacings.

(10) The crystals available for this research were so poor in quality by crystallographic standards that it was impracticable to prepare a three-dimensional Patterson synthesis, or to make critical comparisons between the intensities of the reflexions in different suspension media. Hence many of the conclusions so far reached are tentative; the preparation of larger crystals will be a condition of further progress.

The author wishes to express his gratitude to Sir Lawrence Bragg for his active encouragement throughout the work and for many valuable discussions about the interpretation of the data. He also wishes to thank Dr M. F. Perutz for much help and criticism; Dr W. H. Taylor for his continued interest in the research; Mr M. W. Rees for preparing the crystals of horse myoglobin; and the Department of Scientific and Industrial Research and the Medical Research Council for financial support.

REFERENCES

- Astbury, W. T. & Bell, F. O. 1941 *Nature*, **147**, 696.
 Boyes-Watson, J., Davidson, E. & Perutz, M. F. 1947 *Proc. Roy. Soc. A*, **191**, 83.
 Bragg, W. L. 1949 *Nature*, **164**, 7.
 Buerger, M. J. 1944 *The photography of the reciprocal lattice*. Cambridge, Mass.: ASXRED Monograph No. 1.
 Gutfreund, H. 1950 *Progress in biophysics*, **1** (in the Press). London: Butterworth.
 Keilin, J. & Schmid, K. 1948 *Nature*, **162**, 496.
 Kendrew, J. C. 1948 *Acta Cryst.* **1**, 336.
 Kendrew, J. C. 1949 *Haemoglobin*, p. 149. London: Butterworth.
 Kendrew, J. C. & Perutz, M. F. 1949 *Haemoglobin*, p. 161. London: Butterworth.
 Millikan, G. A. 1939 *Physiol. Rev.* **19**, 503.
 Perutz, M. F. 1949 *Proc. Roy. Soc. A*, **195**, 474.
 Porter, R. R. & Sanger, F. 1948 *Biochem. J.* **42**, 287.
 Roche, J., Derrien, Y. & Vieil, H. 1942 *Bull. Soc. Chim. biol.* **24**, 1016.

- Roche, J. & Vieil, H. 1940 *C.R. Acad. Sci., Paris*, **210**, 314.
Rossi-Fanelli, A. 1940 *Arch. Sci. biol. Napoli*, **26**, 244.
Rossi-Fanelli, A. 1949 *Haemoglobin*, p. 115. London: Butterworth.
Svedberg, T. 1939 *Proc. Roy. Soc. A*, **170**, 40.
Theorell, A. H. T. 1932 *Hoppe-Seyl. Z.* **252**, 1.
Tristram, G. R. 1949 *Haemoglobin*, p. 109. London: Butterworth.
Vieil, H. 1942 *Recherches sur les Myoglobines*. Lons-le-Saunier: Declume.
Wyman, J. 1948 *Advances in protein chemistry*, **4**, 407. New York: Academic Press.
Wyman, J. & Ingalls, E. N. 1943 *J. Biol. Chem.* **147**, 297.
-

The behaviour of supersonic flow past a body of revolution, far from the axis

By G. B. WHITHAM, *Department of Mathematics, University of Manchester*

(Communicated by S. Goldstein, F.R.S.—Received 12 August 1949)

A theory is developed of the supersonic flow past a body of revolution at large distances from the axis, where a linearized approximation is valueless owing to the divergence of the characteristics at infinity. It is used to find the asymptotic forms of the equations of the shocks which are formed from the neighbourhoods of the nose and tail. In the special case of a slender pointed body, the general theory at large distances is used to modify the linearized approximation to give a theory which is uniformly valid at all distances from the axis. The results which are of physical importance are summarized in the conclusion (§ 9) and compared with the results of experimental observations.

1. INTRODUCTION

In the supersonic flow of a compressible fluid past a body of revolution, curved shocks are formed from the neighbourhood of the nose and tail. The characteristics from the surface of the body intersect the shocks, which curve in towards each other and become weaker as a consequence. The present paper describes how the asymptotic shapes and strengths at large distances from the body may be found. However, to do this, a general theory of axisymmetrical flow at large distances from the axis must be developed and it is this topic which receives most attention.

In linearized theory the characteristics are straight and parallel, whereas in fact they are neither. Since the characteristics diverge, the linearized theory becomes less accurate as the distance from the body increases and is, therefore, quite inadequate away from the body. The correct results are established by a new method of attacking non-linear partial differential equations of the hyperbolic type, where a linearized approximation exists and an asymptotic solution with respect to one of the independent variables is required. The same technique may be applied to certain analogous problems of unsteady one-dimensional wave motion, such as the flow behind an expanding spherical shock, as will be described in a later paper. The method

was discovered in collaboration with Mr M. J. Lighthill, who has extended it into a general technique for rendering approximate solutions to physical problems uniformly valid.

The corresponding problem in two dimensions for flow past an aerofoil is comparatively easy, because the details of the exact two-dimensional 'simple-wave' flow of a fluid are known. Lighthill (1944) has shown by a geometrical argument, that the shocks must become parabolae at large distances. Neglecting changes of entropy, Friedrichs (1948) has found an exact expression for the shape of the shock, which gives the asymptotic behaviour deduced by Lighthill. The basis of the method is that the characteristics can be used to connect conditions on the surface of the aerofoil (which are known) with conditions at the shock. This is *only* possible because the characteristics are known to be straight lines with flow variables constant along them. The result, for large y , is

$$x = y \cot \mu_0 - ay^{\frac{1}{2}} - \dots, \quad (1)$$

where a is a constant depending on the aerofoil and μ_0 is the Mach angle of the undisturbed flow. Clearly, such a method cannot be used for the body of revolution, since it would require a detailed knowledge of the characteristics and the flow near the body. However, if only the asymptotic behaviour at large distances is considered, the problem becomes easier.

In any case, the predictions of two-dimensional theory can give approximations to flow past a wing, even of large aspect ratio, only in the region near the surface of the body; its predictions of occurrences at large distances would be wrong. The results for the case of a body of revolution, however, should give an estimate of the truth for flow past any finite body, provided the distance from it is sufficient. Therefore, apart from its scientific interest, the solution may be of practical value; for example, in detecting the passage of a supersonic projectile.

The method of determining the shapes of the shocks requires the asymptotic behaviour of the characteristics. The necessary results are deduced in § 2 by a short unrigorous method, leaving the full discussion and higher approximations until § 5. This presentation is adopted because these first results suggest the approach used later and also they are sufficient for the work of §§ 3 and 4. In § 3, the equation of the front shock is shown to be

$$x = r \cot \mu_0 - br^{\frac{1}{2}} - \dots, \quad (2)$$

using a geometrical result of the shock conditions; and in § 4 some properties of the flow between the shocks are considered. In § 5, the exact equations for irrotational flow are solved using certain expansions of the velocities, for large distances from the axis, in which the form of the characteristics plays an important role. With these full results, further approximations to the results of §§ 3 and 4 are found; and in § 7, this accurate theory which holds at large distances is incorporated with the linearized theory which is valid near the body, for the special case in which the body is slender. In § 8, the form of the rear shock and the flow behind it, topics which have not been discussed in sections mentioned above, are considered, but no complete results can be established. Finally, in § 9 are collected those results which are of importance from a practical point of view.

2. THE CHARACTERISTICS IN AXISYMMETRICAL FLOW

If the velocity of the undisturbed stream is $(U, 0)$ and at a general point (x, r) , in cylindrical co-ordinates, the velocity is $(U + Uu, Uv)$, then the equations of motion for irrotational flow are

$$u_r - v_x = 0, \quad (3)$$

$$\frac{a^2}{U^2} \left(u_x + v_r + \frac{v}{r} \right) - (1 + u)^2 u_x - 2v(1 + u) v_x - v^2 v_r = 0, \quad (4)$$

where a is the local velocity of sound, given by

$$\frac{2a^2}{\gamma - 1} + U^2[(1 + u)^2 + v^2] = \frac{2a_0^2}{\gamma - 1} + U^2, \quad (5)$$

and a_0 is the velocity of sound in the undisturbed stream.

Neglecting terms of the third order in u and v , since these must be small at large distances, the equations become

$$u_r - v_x = 0, \quad (6)$$

$$[\alpha^2 u_x - v_r - v/r] + M^2 u[(\gamma + 1) u_x + (\gamma - 1)(v_r + v/r)] + 2M^2 v v_x = 0, \quad (7)$$

where $M = U/a_0$ and $\alpha = \sqrt{M^2 - 1}$. The linearized theory makes a further approximation by neglecting second-order terms in (7), to give $u_r - v_x = 0$, $\alpha^2 u_x - v_r - v/r = 0$, $\alpha^2 u_{xx} = v_{rx} + v_x/r = u_{rr} + u_r/r$. This equation for u is the equation of cylindrical wave motion, with solution (see Lighthill 1945)

$$u = \int_0^{x-\alpha r} \frac{\lambda(t) dt}{\sqrt{\{(x-t)^2 - \alpha^2 r^2\}}} = \frac{1}{\sqrt{(2\alpha r)}} \int_0^\xi \frac{\lambda(t)}{\sqrt{(\xi-t)}} \left\{ 1 - \frac{\xi-t}{4\alpha r} + \dots \right\} dt \quad (8)$$

$$= \frac{F(\xi)}{r^{\frac{1}{2}}} + \frac{F_1(\xi)}{r^{\frac{3}{2}}} + \dots, \quad (9)$$

say, where $\xi = x - \alpha r$. For large r , $u \simeq F(x - \alpha r) r^{-\frac{1}{2}}$ and from (6), $v \simeq -\alpha u$.

The form of the linearized solution suggests that there is some advantage in using co-ordinates (ξ, r) , where $\xi = x - \alpha r$. With this transformation,

$$\partial/\partial x = \partial/\partial \xi, \quad (\partial/\partial r)_x = (\partial/\partial r)_\xi - \alpha \partial/\partial \xi,$$

so that equations (6) and (7) become

$$u_r - \alpha u_\xi = v_\xi, \quad (10)$$

$$[\alpha^2 u_\xi - v_r + \alpha v_\xi - v/r] + M^2 u[(\gamma + 1) u_\xi + (\gamma - 1)(v_r - \alpha v_\xi + v/r)] + 2M^2 v v_\xi = 0. \quad (11)$$

Substituting for v_ξ from (10), (11) becomes

$$[\alpha u_r - v_r - v/r] + [(\gamma - 1) M^2 (v_r - \alpha u_r + v/r) u + 2M^2 v u_r] + M^2 \{[(\gamma + 1) + (\gamma - 1) \alpha^2] u - 2\alpha v\} u_\xi = 0. \quad (12)$$

In this equation, the coefficient of u_ξ is $O(u)$ as $u \rightarrow 0$, whereas those of u_r and v_r are $O(1)$. Therefore, probably the only quadratic terms which have a marked effect on the solution are those involving u_ξ , since they form the *only* terms in u_ξ . Thus the second

term in (12) is omitted and v is replaced therein by the linearized approximation $-\alpha u$, to obtain

$$2u_r + \frac{u}{r} + ku\xi = 0, \quad (13)$$

where $k = M^4(\gamma + 1)\alpha^{-1}$. Now equation (13) can be written as

$$\frac{\partial}{\partial r}(u\sqrt{r}) + \frac{k(u\sqrt{r})}{2\sqrt{r}} \frac{\partial}{\partial \xi}(u\sqrt{r}) = 0, \quad (14)$$

or as the vanishing of the Jacobian

$$\frac{\partial}{\partial r}(u\sqrt{r}) \frac{\partial}{\partial \xi}(\xi - kur) - \frac{\partial}{\partial \xi}(u\sqrt{r}) \frac{\partial}{\partial r}(\xi - kur) = 0. \quad (15)$$

This implies a functional relation between $u\sqrt{r}$ and $\xi - kur$, so that, if z denotes $-ku\sqrt{r}$, the solution is $u = -z/(kr^{\frac{1}{2}})$, $v = -\alpha u$ and $x = \alpha r - zr^{\frac{1}{2}} - h(z)$, where $h(z)$ is an arbitrary function. In other words, there is a set of curves $z = \text{constant}$, each a parabola with axis in the Mach direction, on each of which $u\sqrt{r}$ is constant. In linearized theory, $u\sqrt{r}$ is constant along characteristics $x - \alpha r = \text{constant}$, which suggests that on the more accurate theory the curves $z = \text{constant}$ are characteristics; that this is so can be shown as follows.

Along a characteristic, $dx/dr = \cot(\mu + \theta)$, where μ is the local Mach angle and θ is the direction of flow. To the first order in u , using $M = \text{cosec } \mu_0$, $\alpha = \cot \mu_0$,

$$\mu = \sin^{-1} \frac{\alpha}{U(1+u)} = \mu_0 - \left(1 + \frac{\gamma-1}{2} M^2\right) \alpha^{-1} u, \quad (16)$$

and $\theta = \tan^{-1}[v/(1+u)] = -\alpha u$, hence

$$\frac{dx}{dr} = \cot(\mu + \theta) = \alpha + \frac{1}{2}ku \quad (17)$$

along a characteristic. But this is also true along $z = \text{constant}$, since

$$dx/dr = \alpha - \frac{1}{2}zr^{-\frac{1}{2}} = \alpha + \frac{1}{2}ku.$$

This completes the results required for §§ 3 and 4: *asymptotically the characteristics are parabolae $x = \alpha r - zr^{\frac{1}{2}} + \text{constant}$, with $u\sqrt{r}$ and $v\sqrt{r}$ constant along them: in fact $u\sqrt{r} = -z/k$, $v\sqrt{r} = \alpha z/k$.*

3. THE SHAPES OF THE SHOCKS

The determination of the shapes of the shocks, as in the two-dimensional case, depends upon the following geometrical property, which is an easy consequence of the shock conditions. *If two regions of supersonic flow are separated by a shock, then, to the first order in the strength, the direction of the shock bisects the Mach directions of the two regions of flow.* (The Mach direction at a point is here understood to mean the outward direction making the local Mach angle with the direction of flow at the point.)

Only bodies of revolution of finite length are considered here, but otherwise the arguments apply quite generally, that is, shell shapes are included and the bow shock

may be detached. The shocks will be formed from the neighbourhoods of the nose and tail, with the characteristics from the body intersecting them. At large distances from the axis, the shocks are weak and approximately at the undisturbed Mach angle, μ_0 , to the stream. Therefore, the asymptotic equation of the front shock is expressed in the form

$$x = r \cot \mu_0 - br^{n+1} = \alpha r - br^{n+1}, \quad (18)$$

where b and n are constants, with $n < 0$ since the shock is weak, and $b > 0$ since the shock is ahead of the Mach cone. The value of n will now be determined.

The characteristics through points on $r = R$ are considered and, in particular, a characteristic C through the point (X, R) which meets the front shock S in (x_1, r_1) , say (figure 1). From the results of § 2 the asymptotic equation of C is

$$x = \alpha r - zr^{\frac{1}{2}} + X - \alpha R + zR^{\frac{1}{2}}. \quad (19)$$

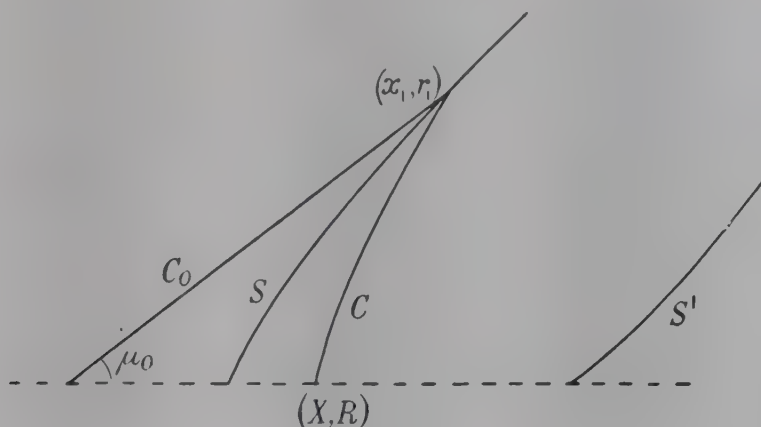


FIGURE 1

The slope of S is given by
$$\frac{dr}{dx} = \frac{1}{\alpha - (n+1)br^n}, \quad (20)$$

and the slope of C by
$$\frac{dr}{dx} = \frac{1}{\alpha - \frac{1}{2}zr^{-\frac{1}{2}}}. \quad (21)$$

At (x_1, r_1) , let the angle of S to the stream be $\mu_0 + \epsilon_1$ and the angle of C $\mu_0 + \epsilon_2$, then, from (20) and (21),

$$\epsilon_1 = b(n+1)r_1^n \sin^2 \mu_0, \quad \epsilon_2 = \frac{1}{2}zr_1^{-\frac{1}{2}} \sin^2 \mu_0. \quad (22)$$

If C_0 is a characteristic of the undisturbed flow ahead of the shock S , the angles of C_0 , S and C , to the stream, are μ_0 , $\mu_0 + \epsilon_1$ and $\mu_0 + \epsilon_2$ respectively. Therefore $\epsilon_2 = 2\epsilon_1$, giving

$$z = 4(n+1)br_1^{n+\frac{1}{2}}. \quad (23)$$

Finally (x_1, r_1) lies on C and S ; hence eliminating x_1 from (18) and (19),

$$(4n+3)br_1^{n+1} - 4(n+1)br_1^{n+\frac{1}{2}}R^{\frac{1}{2}} + \alpha R - X = 0. \quad (24)$$

This equation expresses the condition that the point (x_1, r_1) of the shock and the point (X, R) of the line $r = R$, lie on the same characteristic.

Consider the characteristics meeting the front shock such that $r_1 \gg R$. Such characteristics exist since through any point of the shock there is a characteristic and

it must ultimately meet the line $r = R$. (It cannot meet the rear shock since this must be at an angle to the local flow larger than the local Mach angle.) In this case the second term in (24) can be neglected, if it is assumed that $4n + 3 \neq 0$. *With this assumption*, (24) becomes

$$(4n + 3)br_1^{n+1} = X - \alpha R. \quad (25)$$

But $\epsilon_1 > 0$, since a shock is always at a greater angle than the local Mach angle to the flow ahead, therefore $(n + 1) > 0$. Hence, if $r_1 \rightarrow \infty$ in (25) keeping R fixed, $X \rightarrow \infty$. This is impossible because any characteristic meeting the front shock must always lie between the two shocks. Therefore the assumption $(4n + 3) \neq 0$ is false and hence $n = -\frac{3}{4}$.

Therefore, the equation of the shock is

$$x = \alpha r - br^{\frac{1}{2}}, \quad (26)$$

where b is a constant depending on the body, and the quantity ϵ_1 , which is a measure of the strength of the shock, falls off like $r^{-\frac{1}{2}}$.

If the flow behind the rear shock is assumed to return sufficiently near to the undisturbed state that the perturbation may be neglected in considering a first approximation to the equation of the shock, then the above argument may be applied to show that the equation of the rear shock is

$$x = \alpha r + b_1 r^{\frac{1}{2}}, \quad (27)$$

where b_1 is a constant depending on the body. This will be discussed in detail in § 8; though no conclusive result will be obtained, further grounds will be given for believing that (27) is the correct form. This will be assumed in the next section which would otherwise require modification.

4. PROPERTIES OF THE FLOW BETWEEN THE SHOCKS

From equation (24), the characteristic through the point (X, R) meets the shock in (x_1, r_1) , where

$$br_1^{-\frac{1}{2}} R^{\frac{1}{2}} = \alpha R - X. \quad (28)$$

(Only points (X, R) on characteristics meeting the front shock are considered, since the corresponding results for the rear shock follow in the same way.) From (23), with $n = -\frac{3}{4}$,

$$z = br_1^{-\frac{1}{2}}, \quad (29)$$

therefore along a characteristic meeting the shock at (x_1, r_1) ,

$$u\sqrt{r} = -zk^{-1} = -k^{-1}br_1^{-\frac{1}{2}}, \quad (30)$$

and, using (28), this gives

$$u(X, R) = -k^{-1} \frac{\alpha R - X}{R}, \quad (31)$$

$$v(X, R) = \alpha k^{-1} \frac{\alpha R - X}{R}. \quad (32)$$

These expressions still apply for a point (X, R) on a characteristic meeting the rear shock.

To the first order, the excess pressure at (X, R) between the shocks is given by

$$p = -\rho_0 U^2 u = \rho_0 U^2 k^{-1} \frac{\alpha R - X}{R}, \quad (33)$$

so that the variation of pressure along a line $r = R$ is as shown in figure 2. There is a jump $\rho_0 U^2 k^{-1} b R^{-\frac{1}{2}}$ at the front shock, followed by a linear decrease to $-\rho_0 U^2 k^{-1} b_1 R^{-\frac{1}{2}}$ at the rear shock, after which the form is as yet unknown. The slope of the pressure curve for points between the two shocks is $\rho_0 U^2 / kR$, and this does not depend upon the particular shape of body considered. The result is, in fact, independent of the equations of the shocks, as will be seen in § 6.



FIGURE 2

The aerodynamic forces on a body can be found from the rate of transfer of momentum through a 'control surface' enclosing the body. Since the shapes of the shocks are now known, together with some knowledge of the velocities between them, it may be inquired whether it is possible to calculate the drag. This is not so: *the drag on the body cannot be determined solely from the rate of transfer of momentum to infinity between the shocks.*

The drag is given by

$$D = -2\pi \int_{\Gamma} p r dr - 2\pi \int_{\Gamma} \rho U^2 u (1 + u) r dr + 2\pi \int_{\Gamma} \rho U^2 uv R dx, \quad (34)$$

where the control surface is a cylinder radius R with axis in the direction of the main stream and the integration is along the path Γ in the meridian plane, as shown in figure 3. Since r is constant on AB , the contribution from the region between the two shocks is

$$D_1 = -2\pi \rho_0 U^2 R \int_A^B uv dx \quad (35)$$

$$= \frac{2\pi \rho_0 U^2}{3} \frac{\alpha}{k^2} (b^3 + b_1^3) R^{-\frac{1}{2}}, \quad (36)$$

from (26), (27), (31) and (32). This tends to zero, as R tends to infinity, therefore the above statement follows. The drag, which is not zero, must, therefore, be connected with loss of momentum in the wake due to gain in entropy at the shock wave.

For the two-dimensional case, the same result is found for the drag, although the lift can be determined correctly by the above method.

It seems surprising, since the solution of the flow contains an arbitrary function $h(z)$, depending on the body, that the pressure signature (between the shocks) is the same for all bodies, apart from two arbitrary constants. However, a particular value

of z specifies a particular characteristic so that $h(z)$ gives a measure of the distance between any two characteristics on $r = R$, say. This depends on the body considered, but since it is a finite distance it does not affect asymptotic considerations.

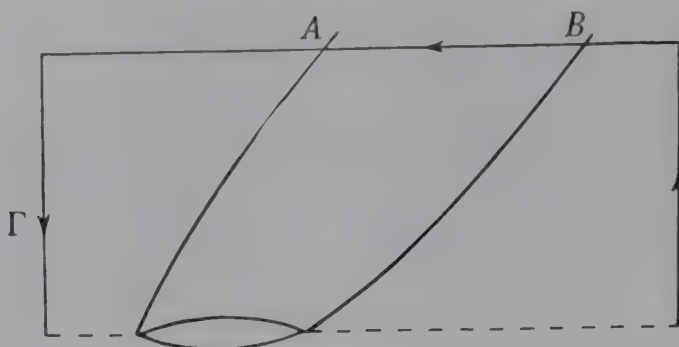


FIGURE 3

5. FULL EXPANSIONS OF THE PHYSICAL QUANTITIES FOR LARGE r

In § 2 certain results were found which open the way for a detailed investigation of axisymmetrical flow, at large distances from the axis. In terms of a variable z the results were $u = -z/(kr^{\frac{1}{2}})$ and $x = \alpha r - zr^{\frac{1}{2}} - h(z)$, but they can be put in a form which compares more directly with linearized theory by taking a new variable y , such that $h(z) = -y$ and $z = -kF(y)$, say. Then

$$u = \frac{F(y)}{r^{\frac{1}{2}}} + \dots, \quad (37)$$

where y is determined by $y = x - \alpha r - kF(y)r^{\frac{1}{2}}$ and $y = \text{constant}$ is a characteristic. The solution given by linearized theory is of the form

$$u = \frac{F(x - \alpha r)}{r^{\frac{1}{2}}} + \frac{F_1(x - \alpha r)}{r^{\frac{3}{2}}} + \dots, \quad (38)$$

$$v = \frac{G(x - \alpha r)}{r^{\frac{1}{2}}} + \frac{G_1(x - \alpha r)}{r^{\frac{3}{2}}} + \dots \quad (39)$$

Comparison of (37) and (38) suggests the substitution of expansions

$$u = \frac{F(y)}{r^{\frac{1}{2}}} + \frac{F_1(y)}{r^{\frac{3}{2}}} + \dots, \quad (40)$$

$$v = \frac{G(y)}{r^{\frac{1}{2}}} + \frac{G_1(y)}{r^{\frac{3}{2}}} + \dots, \quad (41)$$

in the exact equations of motion, from which conditions on y for such solutions to exist and relations between the functions $F, F_1, \dots, G, G_1, \dots$ should be found.

This approach to the solution of non-linear partial differential equations is similar, in some respects, to a method of Poincaré for non-linear ordinary differential equations. For instance, consider the equation for non-linear damped oscillations

$$\frac{d^2x}{dt^2} + \mu f\left(x, \frac{dx}{dt}\right) + x = 0, \quad (42)$$

where μ is a small parameter. When $\mu = 0$, the equation has solution $A \sin t$, periodic with period 2π . To find periodic solutions of (42) for small values of μ , the possibility of a solution $x = A \sin \omega t + O(\mu)$, uniformly in t (i.e. even for large t), is investigated. It is found that this is so only for a certain value of ω depending on μ , and for certain values of the constant A . These restrictions are the conditions of periodicity and take the form

$$\int_0^{2\pi} f(A \sin t, A \cos t) \cos t dt = 0, \quad \omega = 1 - \frac{\mu}{A} \int_0^{2\pi} f(A \sin t, A \cos t) \sin t dt + O(\mu^2). \quad (43)$$

In the same way, (40) and (41) have the form of the linearized solution, and conditions on y are found such that this solution is possible. Later, y is restricted such that $y = \text{constant}$ is a characteristic.

This is the principle of the approach, but in fact it requires modifications. It is found that the expansions (40) and (41) are inadequate and

$$u = \frac{F(y)}{r^{\frac{1}{2}}} + \frac{F_1(y)}{r} + \frac{F_2(y)}{r^{\frac{3}{2}}} + \dots, \quad (44)$$

$$v = \frac{G(y)}{r^{\frac{1}{2}}} + \frac{G_1(y)}{r} + \frac{G_2(y)}{r^{\frac{3}{2}}} + \dots, \quad (45)$$

must be taken. This is reasonable because the presence of u^2 , etc., in the exact equations of motion, immediately introduces terms in r^{-1} . Expansions (44) and (45) are still not quite the correct forms to take, but they will be used since the necessary modifications, which are only slight, can be carried out later. Actually § 2 suggests y is of the form

$$y = x - \alpha r - H(y) r^{\frac{1}{2}} - \dots; \quad (46)$$

in other words x , like u and v , can be expanded in powers of r with coefficients functions of y . Since it is found that the equations of motion can be solved assuming three expansions of this type, it is unnecessary to go through the procedure of finding the behaviour of y *ab initio*.

This introduction shows how the method of solution which will be used is suggested by linearized theory and analogy with Poincaré's method, but in fact it is much more convenient in all this work to use the original variable z instead of y , and write

$$x = \alpha r - z r^{\frac{1}{2}} - m(z) - \dots, \quad (47)$$

$$u = f(z) r^{-\frac{1}{2}} + g(z) r^{-1} + \dots, \quad (48)$$

$$v = -\alpha u + a(z) r^{-1} + b(z) r^{-\frac{3}{2}} + \dots, \quad (49)$$

in place of (44), (45) and (46). To keep the argument general a term in $r^{-\frac{1}{2}}$ should be included in (49), but it would be found to be unnecessary and so is omitted.

The exact equations of steady irrotational flow, by (3), (4) and (5), are

$$u_r - v_x = 0, \quad (50)$$

$$[\alpha^2 u_x - v_r - v/r] + M^2 u[(\gamma + 1) u_x + (\gamma - 1)(v_r + v/r)] + 2M^2 v v_x + M^2 [\frac{1}{2}(\gamma - 1)(u^2 + v^2)(u_x + v_r + v/r) + u^2 u_x + 2uv v_x + v^2 v_r] = 0. \quad (51)$$

From (47), using primes to denote differentiation with respect to z ,

$$z_x = \frac{-1 + \dots}{r^{\frac{1}{2}} + m' + \dots} = -r^{-\frac{1}{2}} + m'r^{-1} + O(r^{-\frac{3}{2}}), \quad (52)$$

$$z_r = \frac{\alpha - \frac{1}{2}zr^{-\frac{1}{2}} - \dots}{r^{\frac{1}{2}} + m' + \dots} = -\alpha z_x - \frac{1}{2}zr^{-1} + \frac{1}{2}zm'r^{-\frac{1}{2}} + O(r^{-2}). \quad (53)$$

From (48) and (49)

$$\begin{aligned} u_x &= \frac{\partial u}{\partial z} z_x = (f'r^{-\frac{1}{2}} + g'r^{-1} + \dots)(-r^{-\frac{1}{2}} + m'r^{-1} + \dots) \\ &= -f'r^{-1} + (f'm' - g')r^{-\frac{1}{2}} + O(r^{-2}), \end{aligned} \quad (54)$$

$$\begin{aligned} v_x &= \frac{\partial v}{\partial z} z_x = -\alpha u_x + (-r^{-\frac{1}{2}} + m'r^{-1} + \dots)(a'r^{-1} + b'r^{-\frac{1}{2}} + \dots) \\ &= -\alpha u_x - a'r^{-\frac{1}{2}} + (a'm' - b')r^{-2} + O(r^{-\frac{3}{2}}), \end{aligned} \quad (55)$$

$$\begin{aligned} u_r &= z_r \frac{\partial u}{\partial z} + \left(\frac{\partial u}{\partial r} \right)_z = -\alpha z_x \frac{\partial u}{\partial z} + (-\frac{1}{2}zr^{-1} + \frac{1}{2}zm'r^{-\frac{1}{2}} + \dots) \\ &\quad \times (f'r^{-\frac{1}{2}} + g'r^{-1} + \dots) - \frac{1}{2}fr^{-\frac{1}{2}} - gr^{-2} + \dots \\ &= -\alpha u_x - \frac{1}{2}(zf' + f)r^{-\frac{1}{2}} + (\frac{1}{2}zm'f' - \frac{1}{2}zg' - g)r^{-2} + O(r^{-\frac{3}{2}}). \end{aligned} \quad (56)$$

It is convenient, before finding v_r , to use equation (50). From (55) and (56), equating coefficients of powers of r ,

$$a' = \frac{1}{2}(zf' + f), \quad (57)$$

$$b' = a'm' - \frac{1}{2}zm'f' + \frac{1}{2}zg' + g = \frac{1}{2}fm' + \frac{1}{2}zg' + g. \quad (58)$$

Hence

$$\begin{aligned} v_r &= -\alpha u_r + \left(\frac{\partial}{\partial r} + z_r \frac{\partial}{\partial z} \right) (ar^{-1} + br^{-\frac{1}{2}} + \dots) \\ &= \alpha^2 u_x + \frac{1}{2}\alpha(zf' + f)r^{-\frac{1}{2}} - \alpha(\frac{1}{2}zm'f' - \frac{1}{2}zg' - g)r^{-2} - ar^{-2} \\ &\quad + (\alpha r^{-\frac{1}{2}} - (\alpha m' + \frac{1}{2}z)r^{-1})(a'r^{-1} + b'r^{-\frac{1}{2}}) + O(r^{-\frac{3}{2}}) \\ &= \alpha^2 u_x + w, \end{aligned} \quad (59)$$

$$\text{where} \quad w = 2\alpha a'r^{-\frac{1}{2}} - [\alpha(zm'f' - zg' - 2g) + a + \frac{1}{2}za']r^{-2} + O(r^{-\frac{3}{2}}), \quad (60)$$

using (58). Substituting in equation (51) and neglecting terms $O(r^{-\frac{1}{2}})$,

$$\begin{aligned} &[-w - (-\alpha fr^{-\frac{1}{2}} - \alpha gr^{-2} + ar^{-2})] + M^2 u[(\gamma + 1)u_x + (\gamma - 1)(\alpha^2 u_x + w - \alpha fr^{-\frac{1}{2}})] \\ &\quad + 2M^2(-\alpha u + ar^{-1})(-\alpha u_x - a'r^{-\frac{1}{2}}) + M^2[\frac{1}{2}(\gamma - 1)(1 + \alpha^2)^2 + 1 + 2\alpha^2 + \alpha^4]u^2 u_x = 0, \end{aligned} \quad (61)$$

$$\begin{aligned} \text{or} \quad &[-w + \alpha fr^{-\frac{1}{2}} + (\alpha g - a)r^{-2}] + (\gamma + 1)M^4 u u_x + (\gamma - 1)M^2(w - \alpha fr^{-\frac{1}{2}})u \\ &\quad - 2M^2 \alpha a u_x r^{-1} + 2M^2 \alpha a' u r^{-\frac{1}{2}} + \frac{1}{2}(\gamma + 1)M^6 u^2 u_x = 0. \end{aligned} \quad (62)$$

The coefficient of $r^{-\frac{1}{2}}$ using (57) gives $zf' + kff' = 0$, where $k = M^4(\gamma + 1)/\alpha$. Hence if $f' \neq 0$ (a hypothesis which will be justified later)

$$f = -z/k. \quad (63)$$

$$\text{From (57) } a' = -z/k, \text{ hence} \quad a = -\frac{z^2}{2k} - A, \quad (64)$$

where A is an arbitrary constant. Next the coefficient of r^{-2} gives that

$$\left[\alpha \left(-\frac{zm'}{k} - zg' - 2g \right) + \alpha g - \frac{z^2}{2k} \right] + \alpha k \left[\frac{g}{k} + \frac{z}{k} \left(\frac{m'}{k} + g' \right) \right] + (\gamma - 1) M^2 \left(-\frac{z}{k} \right) \left(-\frac{2\alpha z}{k} + \frac{\alpha z}{k} \right) + 2M^2 \alpha \left(\frac{z^2}{2k^2} + \frac{A}{k} \right) + \frac{2M^2 \alpha z^2}{k^2} + \frac{M^2 \alpha z^2}{2k^2} = 0, \quad (65)$$

that is
$$\frac{2\alpha M^2 A}{k} + \frac{z^2 M^2}{2\alpha k^2} [M^2(\gamma + 4) - (2\gamma + 5)] = 0. \quad (66)$$

Excluding the trivial result $z = \text{constant}$, this is only true if

$$A = 0 \quad \text{and} \quad M = \sqrt{\{(2\gamma + 5)/(\gamma + 4)\}} = 1.202$$

for air. For other values of M , this contradiction can be eliminated by modifying the original expansions as follows. In the above, let $m(z)$, $b(z)$ stand for

$$m_1(z) \log r + m_2(z), \quad b_1(z) \log r + b_2(z),$$

respectively. ($a(z)$ and $g(z)$ could be assumed to be of this form also, but the equations give $a_1 = 0$ and $g_1 = 0$ making it unnecessary.) This change in m and b only affects terms involving derivatives with respect to r . In (53)

$$z_r = \frac{\alpha - \frac{1}{2}zr^{-\frac{1}{2}} - m_1 r^{-1} - \dots}{r^{\frac{1}{2}} + m' + \dots} = -\alpha z_x - \frac{1}{2}zr^{-1} + \frac{1}{2}zm'r^{-\frac{3}{2}} - m_1 r^{-\frac{3}{2}} + \dots, \quad (67)$$

hence z_r has an extra term $-m_1 r^{-\frac{3}{2}}$. Similarly in (56) u_r has an extra term $-m_1 f' r^{-2}$; in (58) b' has an extra term $m_1 f'$; and in (60) w has an extra term $2\alpha m_1 f' r^{-2}$. Now, from the coefficient of r^{-2} in the equation of motion (62), (66) is replaced by

$$\frac{2\alpha M^2 A}{k} + \frac{z^2 M^2}{2\alpha k^2} [M^2(\gamma + 4) - (2\gamma + 5)] - 2\alpha m_1 f' = 0, \quad (68)$$

hence, from (63),
$$m_1 = -kKz^2 - M^2 A, \quad (69)$$

where
$$K = \frac{M^2}{4\alpha^2 k^2} [M^2(\gamma + 4) - (2\gamma + 5)]. \quad (70)$$

The equations of motion do not determine the functions $g(z)$ and $b(z)$ unless the condition $z = \text{constant}$ is a characteristic is stipulated. The characteristics of the equations

$$\left. \begin{aligned} A_1 u_x + B_1 u_r + C_1 v_x + D_1 v_r + E_1 &= 0, \\ A_2 u_x + B_2 u_r + C_2 v_x + D_2 v_r + E_2 &= 0, \end{aligned} \right\} \quad (71)$$

are given by Courant & Friedrichs (1948):

$$[AC] dr^2 - ([AD] + [BC]) dx dr + [BD] dx^2 = 0, \quad (72)$$

where $[XY]$ denotes $X_1 Y_2 - X_2 Y_1$. For equations (50) and (51),

$$A_1 = D_1 = 0, \quad B_1 = 1, \quad C_1 = -1,$$

and

$$A_2 = \alpha^2 + (\gamma + 1) M^2 u + M^2 \left\{ \frac{1}{2} M^2 (\gamma - 1) + 1 \right\} u^2 + O(r^{-\frac{1}{2}}),$$

$$B_2 = 0,$$

$$C_2 = 2M^2 (-\alpha u + ar^{-1}) (1 + u) + O(r^{-\frac{1}{2}}),$$

$$D_2 = -1 + (\gamma - 1) M^2 u + M^2 \left\{ \frac{1}{2} (\gamma - 1) M^2 + \alpha^2 \right\} u^2 + O(r^{-\frac{1}{2}}).$$

If $z = \text{constant}$ is a characteristic, $dx/dr = \alpha - \frac{1}{2}zr^{-\frac{1}{2}} - m_1r^{-1} + O(r^{-\frac{3}{2}})$; hence substituting in (72),

$$\begin{aligned} & \alpha^2 + (\gamma + 1) M^2 u + M^2 \left\{ \frac{1}{2}(\gamma - 1) M^2 + 1 \right\} u^2 - 2M^2 (-\alpha u - \alpha u^2 + \alpha r^{-1}) (\alpha - \frac{1}{2}zr^{-\frac{1}{2}}) \\ & + \left\{ -1 + (\gamma - 1) M^2 u + M^2 \left(\frac{1}{2}(\gamma - 1) M^2 + \alpha^2 \right) u^2 \right\} \left\{ \alpha^2 - \alpha zr^{-\frac{1}{2}} + \frac{1}{4}z^2r^{-1} - 2\alpha m_1r^{-1} \right\} \\ & + O(r^{-\frac{3}{2}}) = 0, \end{aligned} \quad (73)$$

or

$$\begin{aligned} & (\gamma + 1) M^4 u + \alpha zr^{-\frac{1}{2}} + M^2 \left\{ \frac{1}{2}M^2(\gamma - 1) + 1 \right\} u^2 - 2M^2 \left\{ -\alpha^2 u^2 + \alpha \alpha r^{-1} + \frac{1}{2}\alpha z u r^{-\frac{1}{2}} \right\} \\ & + M^2 \left\{ \frac{1}{2}(\gamma - 1) M^2 + \alpha^2 \right\} \alpha^2 u^2 - (\gamma - 1) M^2 \alpha z u r^{-\frac{1}{2}} - \frac{1}{4}z^2r^{-1} + 2\alpha m_1r^{-1} + O(r^{-\frac{3}{2}}) = 0. \end{aligned} \quad (74)$$

Equating coefficients of $r^{-\frac{1}{2}}$ to zero,

$$(\gamma + 1) M^4 f + \alpha z = 0, \quad (75)$$

which was found previously; this excludes the possibility $f = \text{constant}$ which was noted at equation (63); and

$$\alpha k g + \frac{1}{2}(\gamma + 1) M^6 \frac{z^2}{k^2} + \frac{M^2 \alpha z^2}{k} + 2M^2 \alpha A + \frac{M^2 \alpha z^2}{k} + (\gamma - 1) \frac{M^2 \alpha z^2}{k} - \frac{1}{4}z^2 + 2\alpha m_1 = 0. \quad (76)$$

Hence, using (69),

$$g = K_1 z^2, \quad (77)$$

$$\text{where} \quad K_1 = 2K - \frac{1}{\alpha k} \left[(\gamma + \frac{3}{2}) \frac{M^2 \alpha}{k} - \frac{1}{4} \right] = \frac{M^2}{4\alpha^2 k^2} [M^2(3 - \gamma) - 4]. \quad (78)$$

From (58), with the additional term $m_1 f'$, $b_1(z)$ and $b_2(z)$ are determined in terms of $m_2(z)$, which is arbitrary,

$$b_1 = \frac{1}{3}Kz^3 + B, \quad (79)$$

$$b_2 = -\frac{1}{2k} \int_0^z \zeta m'_2(\zeta) d\zeta + \frac{1}{2}D + \frac{1}{3}(K + 2K_1)z^3 + \frac{M^2 A z}{k}, \quad (80)$$

where B and D are arbitrary constants.

The results are

$$u = -\frac{z}{k}r^{-\frac{1}{2}} + K_1 z^2 r^{-1} + \dots, \quad (81)$$

$$\begin{aligned} v = -\alpha u - \left(\frac{z^2}{2k} + A \right) r^{-1} + \left(\frac{Kz^3}{3} + B \right) r^{-\frac{1}{2}} \log r \\ - \left\{ \frac{1}{2k} \int_0^z \zeta h'(\zeta) d\zeta - \frac{1}{2}D - \frac{1}{3}(K + 2K_1)z^3 - \frac{M^2 A z}{k} \right\} r^{-\frac{1}{2}} + \dots, \end{aligned} \quad (82)$$

$$x = \alpha r - zr^{\frac{1}{2}} + (kKz^2 + M^2 A) \log r - h(z) - \dots, \quad (83)$$

replacing the arbitrary function $m_2(z)$ by $h(z)$.

It is interesting that the possibility of expansions (48) and (49) implies that $z = \text{constant}$ is a characteristic to the second approximation, but not to higher ones; to obtain a unique solution to higher approximations, the nature of these curves must be specified more precisely, as was done above by assuming that they are exactly characteristics.

The velocities u and v were appropriate for solving the equations of motion in this case, but the velocity potential ϕ could have been used and in some problems may be more convenient. The expression for ϕ is easily deduced from the above, since $u = \partial\phi/\partial x$ and $v = \partial\phi/\partial r$; it is

$$\phi = -A \log r - 2B(2 + \log r)r^{-\frac{1}{2}} - C + \frac{z^2}{2k} - \frac{2Kz^3}{3}r^{-\frac{1}{2}}\log r - \left\{ \frac{K_1 z^3}{3} - \frac{1}{k} \int_0^z \xi h'(\xi) d\xi + D \right\} r^{-\frac{1}{2}} + \dots, \quad (84)$$

where C is an arbitrary constant. Actually, in this work, the constants A , B and C are found to be zero.

6. FURTHER TERMS IN THE EQUATION OF THE FRONT SHOCK

In § 3 a first approximation for the equation of the front shock was obtained, but it was assumed to be of the form $x = \alpha r - br^{n+1}$; the possibility of a term $(\log r)^m r^{n+1}$ or other forms was not considered. Now, the arguments are reconsidered making no such assumption, and, in addition, further terms in the equation will be determined.

The equation of a characteristic in the region between the shocks is

$$x = \alpha r - zr^{\frac{1}{2}} + (kKz^2 + M^2A) \log r - h(z) + O(r^{-\frac{1}{2}} \log r), \quad (85)$$

where z is constant, and an error term, of the form expected, has been included for the further terms in the series. For *any* given value of r , x is bounded, since the characteristics must lie between the two shocks, hence z and $h(z)$ are bounded and $h(z)$ may be expanded as $h(0) + zh'(0) + O(z^2)$. Then

$$z = \frac{\alpha r - x + M^2A \log r - h(0)}{r^{\frac{1}{2}} + h'(0)} + O(z^2 r^{-\frac{1}{2}} \log r) + O(r^{-1} \log r). \quad (86)$$

Let the equation of the front shock be $x = \alpha r - f(r)$, then immediately $f'(r) = O(r^{-\frac{1}{2}})$, because the shock must always be at an angle less than the local Mach angle to the local flow behind it. The conditions at the shock give

$$\left(\frac{dx}{dr} - \alpha \right)_s = \frac{1}{2} \left(\frac{dx}{dr} - \alpha \right)_c + O \left[\left(\frac{dx}{dr} - \alpha \right)_c^2 \right], \quad (87)$$

where suffices s and c indicate values on the shock and on the characteristic respectively. Hence

$$-f'(r) = -\frac{1}{4}zr^{-\frac{1}{2}} + \frac{1}{2}M^2Ar^{-1} + O(z^2r^{-1}) + O(r^{-\frac{1}{2}} \log r). \quad (88)$$

Substituting the value of z given by (86), with $x = \alpha r - f(r)$, and using

$$z = O(f'r^{\frac{1}{2}}) + O(r^{-\frac{1}{2}})$$

from (88), a first-order ordinary linear differential equation is obtained for $f(r)$:

$$f(r) - 4r^{\frac{1}{2}}(r^{\frac{1}{2}} + h'(0))f'(r) + M^2A(\log r - 2) - h(0) = O(rf'^2 \log r) + O(r^{-\frac{1}{2}} \log r). \quad (89)$$

This equation is easily solved by successive approximation; since $f'(r) = O(r^{-\frac{1}{2}})$,

(89) gives $f(r) - 4rf'(r) = O(\log r)$ or $d(r^{-\frac{1}{2}}f(r))/dr = O(r^{-\frac{1}{2}} \log r)$

and hence

$$f(r) = br^{\frac{1}{2}} + O(\log r),$$

where b is an arbitrary constant. Using this result, (89) may be written as

$$f(r) - 4rf'(r) = -M^2A(\log r - 2) + h(0) + bh'(0)r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}}\log r), \quad (90)$$

with solution

$$f(r) = br^{\frac{1}{2}} - M^2A(\log r + 2) + h(0) + \frac{1}{2}bh'(0)r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}}\log r). \quad (91)$$

Now, the shock conditions are essentially two in number; if the position of the shock is known, the velocity components u and v are determined behind it. In the above, only one condition has been used, namely, the 'angle property'; as the other, it is convenient to take the condition that the component of momentum tangential to the shock is continuous, that is, $v + u \cot(\mu_0 + \epsilon_1) = 0$, or

$$v + \alpha u = uf'(r). \quad (92)$$

From (88) and (91), $z = br^{-\frac{1}{2}} + O(r^{-\frac{1}{2}})$ at the shock; hence from (81) and (82), equation (92) gives

$$\begin{aligned} -Ar^{-1} + Br^{-\frac{1}{2}}\log r + \frac{1}{2}\left(D - \frac{b^2}{k}\right)r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}}) \\ = \left[-\frac{br^{-\frac{1}{2}}}{k} + O(r^{-1})\right]\left[\frac{1}{4}br^{-\frac{1}{2}} + O(r^{-1})\right], \end{aligned} \quad (93)$$

for all r . Therefore $A = B = 0$ and $b^2 = 2kD$.

The equation of the front shock is

$$x = \alpha r - br^{\frac{1}{2}} - h(0) - \frac{1}{2}bh'(0)r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}}\log r), \quad (94)$$

where $b = \sqrt{(2Dk)}$.

Further approximations to the values of u , v and p , at points between the shocks, found in § 4, are now possible. From (81) and (86), using $x - \alpha r = O(r^{\frac{1}{2}})$,

$$u = \frac{x + h(0) - \alpha r}{kr} - \frac{h'(0)}{k} \frac{x + h(0) - \alpha r}{r^{\frac{1}{2}}} + O(r^{-\frac{1}{2}}\log r), \quad (95)$$

$$p = -\rho_0 U^2 u + O(u^2) = -\frac{\rho_0 U^2}{k} \left\{ \frac{x + h(0) - \alpha r}{r} - h'(0) \frac{x + h(0) - \alpha r}{r^{\frac{1}{2}}} + O(r^{-\frac{1}{2}}\log r) \right\}. \quad (96)$$

These results show that even to this second approximation the pressure falls linearly between the shocks, although the slope of the pressure signature is now

$$\rho_0 U^2 (1 - h'(0)r^{-\frac{1}{2}})/kr,$$

and $h'(0)$ depends upon the particular body under consideration.

7. CONNEXION WITH LINEARIZED THEORY

In this paper a general theory of the flow past a body of revolution has been developed, which is applicable at large distances from the axis. For the special case of a pointed slender body, linearized theory gives a good approximation near the body, and in this section it will be shown how to modify linearized theory to be uniformly valid, for large r as well as small.

For the flow between the shocks, the results of § 5 written in terms of the variable $y = -h(z)$, with $z = -kF(y)$ say, are

$$u = \frac{F(y)}{r^{\frac{1}{2}}} + \frac{k^2 K_1 F^2(y)}{r} + \dots, \quad (97)$$

$$v = -\alpha u - \frac{\frac{1}{2}kF^2(y)}{r} - \frac{\frac{1}{3}k^3KF^3(y)\log r}{r^{\frac{3}{2}}} - \frac{\frac{1}{2}\int_{y_0}^y F(y') dy' - \frac{1}{2}D + \frac{1}{3}k^3(K + 2K_1)F^3(y) + \dots}{r^{\frac{3}{2}}}, \quad (98)$$

where $y = x - \alpha r - kF(y)r^{\frac{1}{2}} - k^3KF^2(y)\log r + \dots$, (99)

and y_0 is the zero of $F(y)$, that is, the value of y on the 'dividing' characteristic C^* (which is asymptotic to the straight line $x = \alpha r + y_0$). Linearized theory has

$$\phi = -\frac{1}{2\pi} \int_0^{x-\alpha r} \frac{S'(t) dt}{\sqrt{\{(x-t)^2 - \alpha^2 r^2\}}}, \quad (100)$$

where $S(x)$ is the cross-sectional area of the body at a distance x along the axis from the nose. For large r , this gives

$$u = \frac{F(\xi)}{r^{\frac{1}{2}}} + O\left(\frac{1}{r^{\frac{3}{2}}}\right), \quad (101)$$

$$v = -\alpha u - \frac{\frac{1}{2}\int_0^\xi F(\xi') d\xi'}{r^{\frac{3}{2}}} + O\left(\frac{1}{r^{\frac{5}{2}}}\right), \quad (102)$$

where $\xi = x - \alpha r$ and $F(\xi) = -\frac{1}{2\pi\sqrt{(2\alpha)}} \int_0^\xi \frac{S''(t) dt}{\sqrt{(\xi-t)}}. \quad (103)$

The above results show that linearized theory would have the correct *form* of expansions for u and v at large distances, with only terms in F^2 , F^3 , ... neglected, if the arbitrary constant D had the value $-\int_0^{y_0} F(y') dy'$. For this case of a slender body, F is small, thus the omission of F^2 , F^3 , ... causes only slight errors; the failure of linearized theory at large distances is due to the second modification which it makes to (97) and (98), that is, the replacement of y by $\xi = x - \alpha r$. This suggests that linearized theory will give a good approximation everywhere provided $x - \alpha r$ is replaced therein by y , where $y = x - \alpha r - kF(y)r^{\frac{1}{2}}$; further, D is determined in terms of $F(y)$ to give

$$b^2 = 2kD = -2k \int_0^{y_0} F(y') dy'. \quad (104)$$

From (103), $F(y)$ has a factor δ^2 , where δ is the fineness ratio for the body considered, hence near the body $y = x - \alpha r + O(\delta^2)$, and the procedure suggested above has, from (100),

$$\phi = -\frac{1}{2\pi} \int_0^y \frac{S'(t) dt}{\sqrt{\{(y-t)(y+2\alpha r-t)\}}} \quad (105)$$

$$= -\frac{1}{2\pi} \int_0^{x-\alpha r} \frac{S'(t) dt}{\sqrt{\{(x-t)^2 - \alpha^2 r^2\}}} + O(\delta^4). \quad (106)$$

But linearized theory already neglects terms $O(\delta^4)$, so that the modified theory is just as accurate near the body as well as giving a valid first approximation for large r .

The dependence of the various constants, occurring in § 6, upon the Mach number of the undisturbed flow and the fineness ratio of the body may be found from (103) and (104). Since $h(z) = -y$ and $z = -kF(y)$, for any particular shape of body

$$b \propto M^2(M^2 - 1)^{-\frac{1}{2}} \delta, \quad h'(0) = -(dy/dz)_{z=0} = \{kF'(y_0)\}^{-1} \propto M^{-4}(M^2 - 1)^{\frac{1}{2}} \delta^{-2}$$

and $h(0) = -y_0$ is independent of M and δ . The strength of the shock, measured by $\alpha - (dx/dr)$, is

$$\frac{1}{4}br^{-\frac{1}{2}} - \frac{1}{8}bh'(0)r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}} \log r),$$

and the strength at the nose (where it is maximum) is found from the theory of flow past a cone to be $O(\delta^4)$, hence the prediction of the shape of the shock is only applicable at distances of order δ^{-4} .

Example. This modified theory, valid for all r , is now discussed for the special case of the symmetrical parabolic profile $r = 2\delta(x - x^2)$, $0 \leq x \leq 1$; the body is obtained by revolving an arc of a parabola about a line perpendicular to its axis. In this case, $S''(t) = 8\pi\delta^2(1 - 6t + 6t^2)$ and

$$F(y) = -\frac{4\sqrt{2}}{5\sqrt{\alpha}} \delta^2 y^{\frac{1}{2}} (5 - 20y + 16y^2), \quad (107)$$

$$F'(y) = -\frac{2\sqrt{2}}{\sqrt{\alpha}} \delta^2 y^{-\frac{1}{2}} (1 - 12y + 16y^2). \quad (108)$$

Hence $F(y) = 0$ at $y = 0$, $(20 \pm \sqrt{80})/32$ and the form of $F(y)$ is as shown (figure 4). The region which applies to the flow at large distances must be AC , since earlier considerations have shown that, far from the axis, $F(y)$ increases nearly linearly from negative to positive values, and therefore y_0 is the zero of $F(y)$ at which $F'(y)$ is positive. The values of the various constants are

$$b = 0.91 M^2 \alpha^{-\frac{1}{2}} \delta, \quad h(0) = -0.35 \quad \text{and} \quad h'(0) = 0.071 M^{-4} \alpha^{\frac{1}{2}} \delta^{-2}.$$

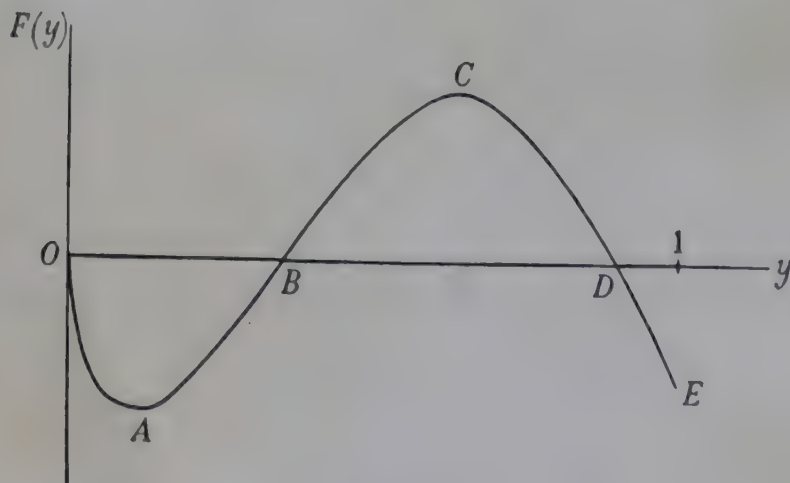


FIGURE 4

The second zero of $F(y)$ at D suggests immediately that a shock is formed at the rear before the tail is reached because two straight characteristics in the flow could not be accounted for; there is also the possibility of the characteristics corresponding to the regions CE and OA forming envelopes. An envelope of characteristics is given, in terms of a parameter y , by

$$kF'(y) R^{\frac{1}{2}} + 1 = 0, \quad (109)$$

$$X = \alpha R + kF(y) R^{\frac{1}{2}} + y. \quad (110)$$

Since the positive root is always taken for $R^{\frac{1}{2}}$, the characteristics actually form an envelope when $F'(y) < 0$, that is, for OA and CE . Equation (110) shows that if $F(y) < 0$, that is, for DE , the point (X, R) always lies ahead of the undisturbed Mach cone from the corresponding point on the body. Hence, since the envelope cannot appear in the flow, a shock must be formed before the undisturbed Mach cone from D , at a distance $O(\delta^{-4})$ from the body.

The appearance of a third shock as above or due to irregularities in the shape of the body (as seen in photographs of bullets in flight) may be discussed by the methods developed in this paper. It is found that such a shock either joins with the rear shock or if three shocks persist to infinity, the middle one has the equation

$$x = \alpha r + c_0 + d_0 r^{-\frac{1}{2}} + O(r^{-1} \log r),$$

where c_0 and d_0 are constants depending on the body. In the latter case, the pressure falls linearly from positive to negative values between any two shocks, and this could not be the case for the example considered; the third shock joins with the rear one.

* In OA , $F(y) < 0$, therefore the front envelope is a distance $O(R^{\frac{1}{2}})$ ahead of the undisturbed Mach cone, whereas the shock is a distance $O(R^{\frac{1}{2}})$ ahead. Therefore the front envelope does not occur (the characteristics meet the shock before meeting one another).

8. THE REAR SHOCK WAVE

In the previous sections, discussions of the rear shock and the flow downstream of it have, in general, been avoided; the results do not follow simply from suitable modifications of the arguments employed at the front shock. In fact, the author cannot give complete results, although some definite answers are found which give support to his conclusions.

The flow behind the rear shock is axisymmetrical, and hence, with the assumption of irrotational motion a solution of the equations for u and v is given in § 5 in terms of a variable z_1 , an arbitrary function $h_1(z_1)$ and certain arbitrary constants $A_1, B_1, D_1, \dots$. However, for this region, $h_1(z_1)$ is not necessarily bounded, and since repeated integrals of z_1 with respect to h_1 occur in the solution, the expansions are found to be valid only if h_1/r , that is, $(x - \alpha r)/r$, is bounded. At the shock where h_1 is maximum, it is at most $O(r^{\frac{1}{2}})$, hence the expansions will be valid in a certain neighbourhood of the shock.

Consider characteristics C and C_1 through a point of the shock, which belong to the regions ahead of and behind it respectively. Then C is of the form

$$x = \alpha r - zr^{\frac{1}{2}} - h(z) + O(z^2 \log r), \quad (111)$$

C_1 is of the form

$$x = \alpha r - z_1 r^{\frac{1}{2}} + M^2 A_1 \log r - h_1(z_1) + O(z_1^2 \log r), \quad (112)$$

and the equation of the shock is taken to be

$$x = \alpha r + f_1(r). \quad (113)$$

The angle property at the shock gives

$$z + z_1 = -4r^{\frac{1}{2}}f_1'(r) + 2M^2 A_1 r^{-\frac{1}{2}} + O(z^2 r^{-\frac{1}{2}}) + O(z_1^2 r^{-\frac{1}{2}}), \quad (114)$$

and from (111) and (113)

$$z = -\frac{f_1(r) + h(0)}{r^{\frac{1}{2}} + h'(0)} + O(z^2 r^{-\frac{1}{2}} \log r), \quad (115)$$

$$\text{hence} \quad z_1 = \frac{f_1(r) + h(0)}{r^{\frac{1}{2}} + h'(0)} - 4r^{\frac{1}{2}}f_1'(r) + 2M^2 A_1 r^{-\frac{1}{2}} + O(z^2 r^{-\frac{1}{2}} \log r) + O(z_1^2 r^{-\frac{1}{2}}). \quad (116)$$

Equations (115) and (116) are the basis for the conclusion that the rear shock is

$$x - \alpha r = b_1 r^{\frac{1}{2}} + \text{smaller terms}, \quad (117)$$

where b_1 is a constant depending on the body. The reason is that (117) is the only form of $f_1(r)$ which makes z_1 of appreciably smaller order than z , that is, which makes the perturbation velocities and pressure behind the rear shock appreciably smaller than between the shocks. For, from (115), $z = -b_1 r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}})$ and, from (116), $z_1 = O(r^{-\frac{1}{2}})$. This 'small' perturbation behind the shock is expected from comparison with the two-dimensional problem and also the pressure signature obtained from experiments on the flight of bullets show this result. The expression 'appreciably smaller' was used because, for example, $f_1(r) = O(r^{\frac{1}{2}} \log^{\frac{1}{2}} r)$ (which arises as a possibility) gives

$$z = O(r^{-\frac{1}{2}} \log^{\frac{1}{2}} r), \quad z_1 = O(r^{-\frac{1}{2}} \log^{-\frac{1}{2}} r)$$

but is not considered to exhibit the required behaviour.

At the shock, $x - \alpha r = b_1 r^{\frac{1}{2}}$ and z_1 is at most $O(r^{-\frac{1}{2}})$, therefore, from (112) $h_1 \sim -b_1 r^{\frac{1}{2}}$ and tends to $-\infty$ as $z_1 \rightarrow 0$. The characteristics tend to infinity downstream like $r^{\frac{1}{2}}$, where r is the distance from the axis at which they meet the shock, and become straighter since $z_1 \rightarrow 0$; thus the whole field behind the shock influences its shape. This inability to give a definite form for the equation of the shock is not due to the conditions being insufficient to determine it but to the difficulty of completing a solution over the whole field. The arbitrary functions $h(z)$ and $h_1(z_1)$ are treated as known, essentially being determined from the flow near the axis, but for the rear shock, in contrast to the case of the front shock, the precise form of $h_1(z_1)$ must be known before any approximation to the equation can be given.

However, in the special case of a slender body, pointed at both ends, the form of the arbitrary function may be obtained from the modification of the linearized results for this region by substitution of y_1 , determined by $y_1 = x - \alpha r - kF_1(y_1)r^{\frac{1}{2}}$, for $\xi = x - \alpha r$. Linearized theory has

$$u = -\frac{1}{2\pi} \int_0^1 \frac{S''(t) dt}{\sqrt{\{(\xi - t)(\xi - t + 2\alpha r)\}}} \quad (118)$$

for $\xi > 1$, which, when expanded as before for large r , gives

$$u = -\frac{1}{2\pi\sqrt{(2\alpha r)}} \int_0^1 \frac{S''(t) dt}{\sqrt{(\xi-t)}} + \frac{1}{8\pi\alpha\sqrt{(2\alpha)} r^{\frac{1}{2}}} \int_0^1 S''(t) \sqrt{(\xi-t)} dt + \dots$$

For the flow at large distances from the axis ξ is large, hence this expansion will be valid above some line $x - \alpha r = Kr$. Replacing ξ by y_1 ,

$$F_1(y_1) = -\frac{1}{2\pi\sqrt{(2\alpha)}} \int_0^1 \frac{S''(t) dt}{\sqrt{(y_1-t)}}, \quad (119)$$

and it is found from the expansion for v that $A_1 = B_1 = D_1 = 0$. Expanding $F_1(y_1)$ for large y_1 , using $S(0) = S(1) = S'(0) = S'(1) = 0$,

$$F_1(y_1) = -\frac{V}{y_1^{\frac{3}{2}}} + O(y_1^{-\frac{5}{2}}), \quad (120)$$

where
$$V = \frac{3}{8\pi\sqrt{(2\alpha)}} \int_0^1 S(t) dt. \quad (121)$$

In the previous notation this means $h_1(z_1) = -(Vk/z_1)^{\frac{2}{3}} + O(1)$ as $z_1 \rightarrow 0$.

Now, the condition for the tangential component of momentum to be continuous through the shock is

$$z_1^2 - z^2 + 2(z_1 - z) r^{\frac{1}{2}} f_1'(r) + kD r^{-\frac{1}{2}} + O(z_1^{\frac{2}{3}} r^{-\frac{1}{2}}) + O(z r^{-\frac{1}{2}}) = 0 \quad (122)$$

which, with (115) and (116), gives $f_1(r) = br^{\frac{1}{2}} + O(1)$. This shows that at the shock $h_1 \sim -br^{\frac{1}{2}}$, hence $z_1 \sim Vkb^{-\frac{2}{3}} r^{-\frac{5}{6}}$ and from (116), with $A_1 = 0$,

$$f_1(r) = br^{\frac{1}{2}} = h(0) + \frac{2Vk}{3b^{\frac{2}{3}}} r^{-\frac{1}{2}} + \dots \quad (123)$$

From the equation of a characteristic,

$$x = \alpha r - z_1 r^{\frac{1}{2}} + (Vk/z_1)^{\frac{2}{3}} + O(1)$$

and the maximum value of z_1 is $O(r^{-\frac{5}{6}})$, hence

$$z_1 = Vk(x - \alpha r)^{-\frac{3}{2}} [1 + O(r^{-\frac{1}{2}})]$$

and

$$u = -\frac{V}{r^{\frac{1}{2}}(x - \alpha r)^{\frac{3}{2}}} + O(r^{-\frac{5}{2}}), \quad (124)$$

$$p = \frac{\rho_0 U^2 V}{r^{\frac{1}{2}}(x - \alpha r)^{\frac{3}{2}}} + O(r^{-\frac{5}{2}}). \quad (125)$$

Thus the indication of slender body theory, which may be taken as a guide (if an uncertain one) in other cases, is that the equation of the rear shock is (123), with the same b and $h(0)$ as in (94); the pressure directly behind the rear shock is $O(r^{-\frac{5}{2}})$ as against $O(r^{-\frac{1}{2}})$ ahead of it; and it asymptotes to zero from this value as indicated in (125).

9. CONCLUSION

In this section those results that are important from a practical point of view are collected together.

For a general body of revolution of finite length, the equations of the front and rear shocks are $x = \alpha r - br^{\frac{1}{2}} - c - dr^{-\frac{1}{2}} - \dots$, $x = \alpha r + b_1 r^{\frac{1}{2}} + \dots$, respectively, for large r ,

where $\alpha = \sqrt{M^2 - 1}$. For constant r , the pressure falls approximately *linearly* between the shocks, the slope of the pressure signature being

$$-0.58 M^{-2} \alpha r^{-1} \{1 - 2b^{-1} dr^{-\frac{1}{2}} + O(r^{-\frac{1}{2}} \log r)\} \text{ atmospheres/unit distance,} \quad (126)$$

where γ is taken as 1.4 and the length of the body is taken as the unit of distance. It is important to note that, to a first approximation, this is independent of the particular body considered, depending only on the Mach number M and the distance r from the axis.

For slender bodies, details of the constants may be given. The symmetrical parabolic profile of fineness ratio δ will be taken as typical of bodies pointed at both ends, and the body with profile $r = \frac{1}{2} \delta x \sqrt{3 - 2x}$, $0 \leq x \leq 1$, typical of shell-shapes, although the dependence on M and δ is the same for all slender bodies.

In the first case,

$$b = b_1 = 0.91 M^2 \alpha^{-\frac{1}{2}} \delta, \quad c = -0.35, \quad d = 0.032 M^{-2} \alpha^{\frac{1}{2}} \delta^{-1}.$$

At a distance r , the pressure jump at the front shock is $0.53 \alpha^{\frac{1}{2}} \delta r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}})$ atmospheres followed by a linear fall at a rate $0.58 M^{-2} \alpha r^{-1} + O(r^{-\frac{1}{2}})$ atmospheres/unit distance to the value $-0.53 \alpha^{\frac{1}{2}} \delta r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}})$ atmospheres at the rear shock when it jumps to $0.063 M^{-3} \alpha^{\frac{3}{2}} \delta^{-\frac{1}{2}} r^{-\frac{1}{2}}$ atmospheres and asymptotes to zero as shown by (125).

In the second case,

$$b = 0.94 M^2 \alpha^{-\frac{1}{2}} \delta, \quad c = -0.75, \quad d = 0.16 M^{-2} \alpha^{\frac{1}{2}} \delta^{-1}.$$

The pressure jump at the front shock is $0.55 \alpha^{\frac{1}{2}} \delta r^{-\frac{1}{2}} + O(r^{-\frac{1}{2}})$ atmospheres with the same rate of fall as before between the shocks, to a value of order $r^{-\frac{1}{2}}$ at the rear shock. It is expected to return to a small value compared with this behind the rear shock, but the exact form is not known for a shell shape.

If a third shock is formed between the two principal ones, it either meets the rear shock (as in the case of the symmetrical parabolic profile) or is asymptotic to a straight line and its strength is $O(r^{-\frac{1}{2}})$; the pressure falls linearly between any two shocks at the same rate as above.

Du Mond, Cohen, Panofsky & Deeds (1946) have given experimental data in good agreement with these predictions; measurements made at various distances r from a bullet moving with supersonic speed, give the distance between the two shocks to be proportional to r^n , where n differs from 0.25 by not more than 0.04 for the four sizes of bullet used. Further, from (126), the product of the slope of the pressure signature and the distance r is a constant $0.58 \sqrt{(M^2 - 1)}/M^2 = 0.252$ for $M = 2$; the experimental value is 0.238. The authors also suggest a theoretical explanation of these observed results, but many unproved assumptions are made and they emphasize that in order to account for the $-\frac{3}{4}$ power law of amplitude it is very definitely necessary to take into account the degradation of energy into heat as the wave propagates; the present paper gives an explanation without such considerations.

The author wishes to thank Mr M. J. Lighthill for his continued help and encouragement in the solution of this problem; also Professor Goldstein for the keen interest he has shown, and the advice he has given.

REFERENCES

- Courant, R. & Friedrichs, K. O. 1948 *Supersonic flow and shock waves*, p. 41. New York: Interscience Publishers.
- Du Mond, J. W. M., Cohen, E. R., Panofsky, W. K. H., & Deeds, E. 1946 A determination of the wave forms and laws of propagation and dissipation of ballistic shock-waves. *J. Acoust. Soc. Amer.* **18**, 97–118.
- Friedrichs, K. O. 1948 Formation and decay of shock-waves. *Commun. Appl. Math.* **1**, no. 3.
- Lighthill, M. J. 1944 The conditions behind the trailing edge of the supersonic aerofoil. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1930.
- Lighthill, M. J. 1945 Supersonic flow past bodies of revolution. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 2003.

Surface deformation and friction of metals at light loads

BY J. R. WHITEHEAD

*Research Laboratory on the Physics and Chemistry of Rubbing Solids,
Department of Physical Chemistry, University of Cambridge*

(Communicated by F. P. Bowden, F.R.S.—Received 23 August 1949—
Read 3 November 1949)

[Plates 2 to 6]

A study has been made of the interaction of sliding metals when the normal load between them is very small. A new apparatus is described which enables friction to be measured, under controlled conditions of sliding, down to loads of a few milligrams. The deformation of the sliding surfaces has been studied by light and electron microscopy. These techniques have been used to investigate the validity of Amontons's law (proportionality of frictional force to load) over a very wide range of loads, from a few milligrams to several kilograms.

Experimental results on electrolytically polished copper reveal a departure from Amontons's law when the normal load is less than a few grams. A corresponding change is observed in the deformation within the track of sliding. At light loads the friction is low and the damage is slight. This is attributed to the effect of the thin film of oxide formed by contact with the atmosphere. A thicker film of oxide, formed by heating in air, reduces the friction at heavy loads. Results of observations on copper surfaces prepared in various ways show that the influence of surface roughness upon friction is not great.

The results of further experiments show that the friction of silver on silver and of aluminium on aluminium is constant and Amontons's law holds over the whole range of loads. It appears that the thin film of oxide on aluminium is penetrated, even at the lightest loads. The results are discussed in relation to those for copper, and the difference in behaviour of the various metals is attributed to the contrasting properties of their oxides.

Isolated experiments on sapphire and diamond show that the coefficient of friction is low and constant for these non-metals, and a few experiments made on boundary lubricated metals are recorded.

In spite of the deviations from Amontons's law on some metals, the results of all the experiments on dry sliding emphasize the essential similarity of the frictional process and of the basic mechanism over an enormous range of loads.

INTRODUCTION

Recent work on the friction of metals and on the mechanism of sliding has shown that the observed phenomena may be explained in terms of the formation and

subsequent breaking of metallic junctions. For instance, Bowden & Leben (1939) have described the processes of metallic friction in terms of the welding of the two metals at local points of contact, and later work (see Bowden & Tabor 1946) has supported this view. Since the most carefully prepared surfaces contain irregularities which are large compared with molecular dimensions, the real area of contact between the surfaces will be very small and the local pressure will, in general, correspond to the yield point of the metal (Bowden & Tabor 1939; Holm 1946). This intense pressure will cause the formation of localized welded junctions at the points of contact. The frictional resistance is largely due to these, but is also due in some part to the displacement of metal from the path of the slider. The relative magnitudes of the 'shearing' and 'ploughing' terms depend on the relative mechanical properties of the two surfaces, and under many conditions of sliding the 'ploughing' term is small.

It has been shown that the real area of contact is almost independent of the apparent area of the surfaces and is determined primarily by the applied load (Bowden & Tabor 1943). This explains the two classical laws which state that friction is independent of the area of the surfaces and is directly proportional to the applied load (Amontons's law).

Moore (1948) has provided clear visual evidence of the welding process in his photomicrographs of taper sections of the damaged surfaces. It is also clear that even under conditions of good boundary lubrication a localized welding may occur through the lubricant film.

Most of the previous work on the mechanism of friction and surface deformation of metals has been carried out with relatively heavy loads between the sliding surfaces. The purpose of the present work is to investigate the interaction between metals down to very light loads, where the normal force is only a few milligrams and to study the surface deformation under these conditions. It was expected that the influence of thin films of foreign material such as oxide or a deliberately applied boundary lubricant would have an increased effect at the lighter loads. This was, indeed, found to be so.

The surface deformation at light loads was so minute that it was necessary to use the electron microscope in conjunction with the light microscope to study it. It will be shown that the mechanism of sliding is strikingly the same at light and heavy loads. For certain metals such as aluminium and silver Amontons's law holds from the heaviest down to the lightest loads. For other metals such as copper a change occurs when the load is reduced below a certain value. The coefficient of friction is much smaller and the nature of the deformation alters. It is evident that the thin film of oxide present on these metals at room temperature is behaving in a protective manner at light loads, and the friction at loads below the transition corresponds to the sliding of oxide on oxide. The higher friction corresponds to a penetration and breakdown of the oxide film.

There is additional evidence that the very thin adherent oxide layer, formed at room temperature, behaves differently from a thick oxide layer deliberately built up by heating in air. It also appears that the roughness of the metal surface has comparatively little effect on the friction.

Preliminary experiments on the boundary lubrication of electrolytically polished metal surfaces by long-chain fatty acids show that Amontons's law breaks down at light loads and the coefficient of friction *rises*.

In spite of anomalies due to the influence of surface films, the results of all the experiments described emphasize the essential similarity of the frictional process over an enormous range of loads.

EXPERIMENTAL

The experiments were designed to measure the friction between a hemispherical upper slider and a plane lower surface during controlled sliding at a low speed (0.01 cm./sec.). The use of a hemispherical slider facilitated the study of the deformation of the lower surface by confining the damage to a narrow track.

Measurements of friction were made at a number of different values of normal load, ranging from 0.010 to 10,000 g. Experiments by previous workers have provided a large number of results relating to the range 200 to 10,000 g. The emphasis of the present work is on measurements at loads below 200 g.

(1) Apparatus

The arrangement of the apparatus for measuring friction is shown in figure 1. The lower specimen was clamped to a turn-table which was rotated slowly by a synchronous electric clock motor. Metal specimens were machined to the shape of a truncated cone and were clamped to the turn-table by means of a brass ring having a centre hole shaped to conform to the pitch of the cone. Both specimen and turn-table were lapped flat so that the sliding surface moved accurately parallel to the bed of the apparatus. Figure 2, plate 2, shows the appearance of the mounted specimen. The angular velocity of the turn-table was $4.5^\circ/\text{min}$. This made the speed of sliding rather less than 0.01 cm./sec. at the mean radius of 6.67 cm.

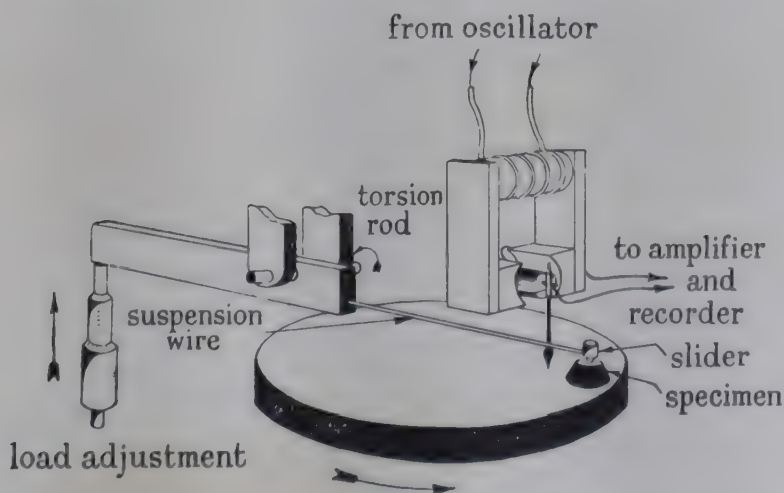


FIGURE 1. Sketch of essential parts of friction apparatus.

The upper surface was formed on the tip of a slider machined out of the desired material. It was approximately hemispherical in shape with a tip radius of 0.4 mm. The slider was held in a boss carried on the end of a spring steel suspension wire.

The suspension wire was attached to a steel beam which was pivoted on a torsion bar and had its remote end supported on the head of a metric micrometer. The lever arm ratio was unity. The torsion bar served to hold the steel beam in contact with the micrometer head.

A normal load was applied to the slider by raising the remote end of the beam. After the two surfaces came into contact a further displacement caused the suspension wire to bend under a known deflexion. The restoring force in the suspension wire loaded the slider against the lower surface. The value of load corresponding to a given deflexion was calibrated by direct comparison with standard weights. Six interchangeable wires covered the load range from 0.010 to 100 g.

The frictional force during sliding caused a lateral deflexion of the slider assembly. This continued until the restoring force in the wire was sufficient to overcome the friction. The lateral deflexion of the wire, representing the frictional force, was measured by means of an electromagnetic pick-up and amplifier and was recorded directly on the paper trace of an electromagnetic pen recorder. Figure 3 shows a typical friction trace relating to steel sliding on aluminium.

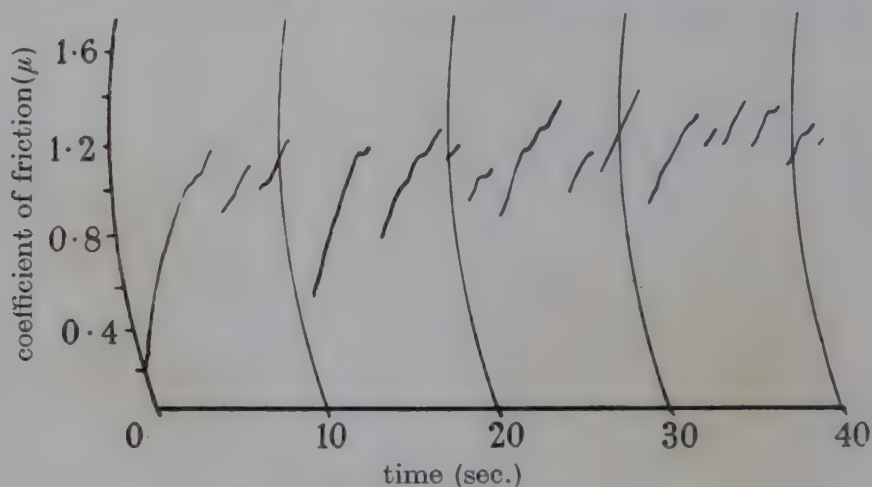


FIGURE 3. Typical friction record (tracing): steel on aluminium.

Various precautions were taken to ensure that the restoring force was applied along the track of sliding and to minimize errors arising from torsion and asymmetry in the suspension. The estimated maximum error in the measurement of friction was $\pm 5\%$.

A few additional measurements of friction were carried out at loads between 200 and 10,000 g., using an apparatus described elsewhere by Bowden & Leben (1939).

(2) *Preparation of specimens*

Each metallic specimen was machined to shape with the upper and lower surfaces accurately parallel. It was then annealed and roughly degreased by washing in an organic solvent, followed by a few minutes in a still containing acetone and isopropyl ether in equal parts. All operations subsequent to degreasing were carried out by handling the specimen with degreased instruments.

The specimen was gently abraded under water on a succession of carborundum papers of decreasing roughness, ending with 600 grade which left irregularities on the surface of the order of 0.1μ in height.

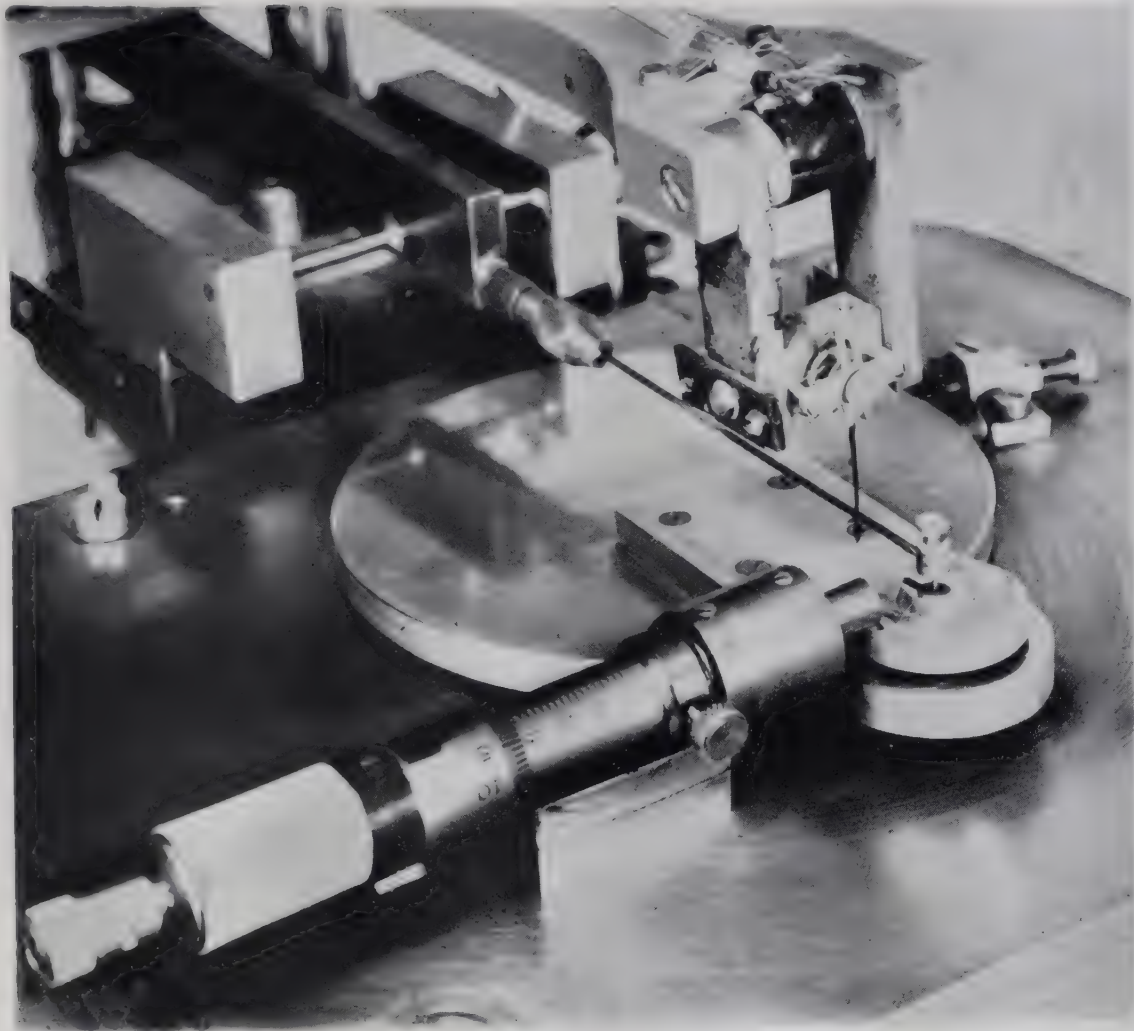
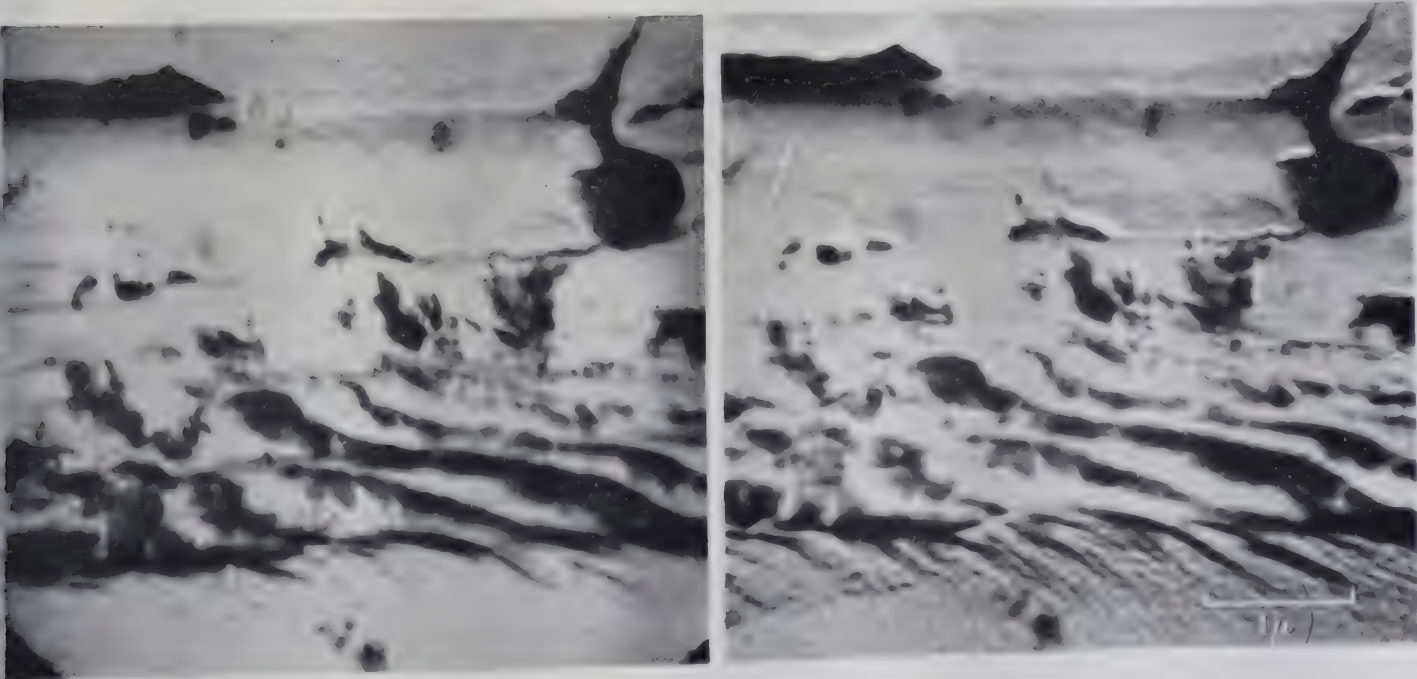


FIGURE 2

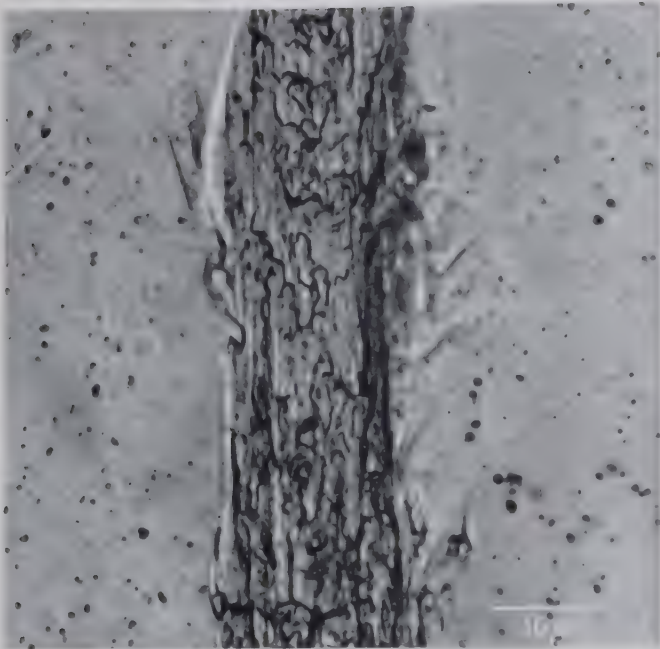


a

b

FIGURE 4

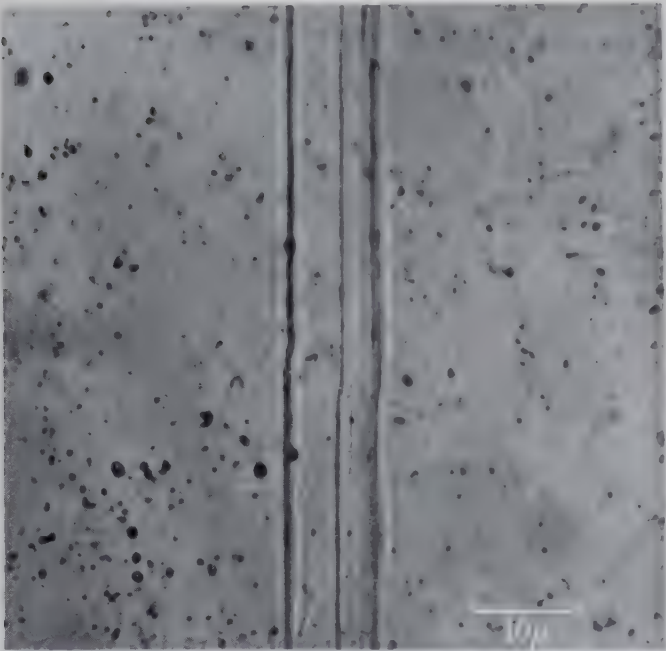
(Facing p. 112)



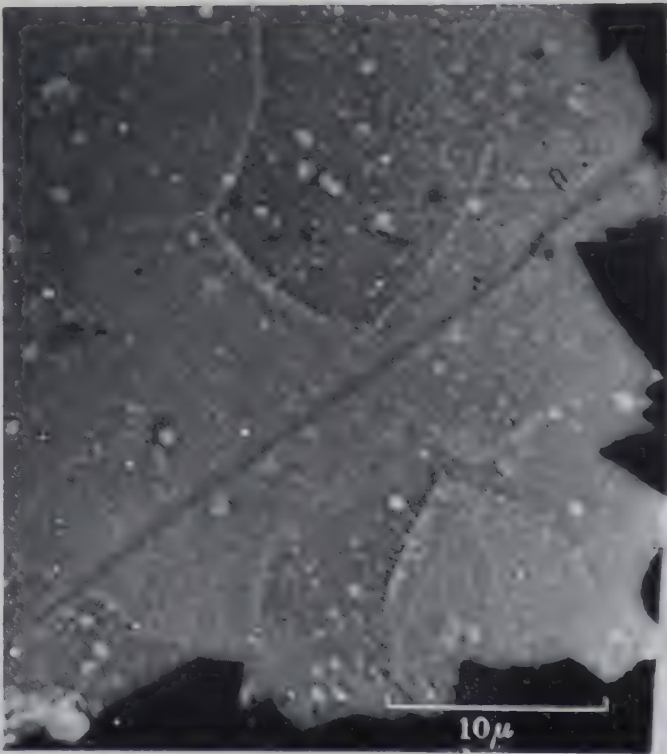
6a



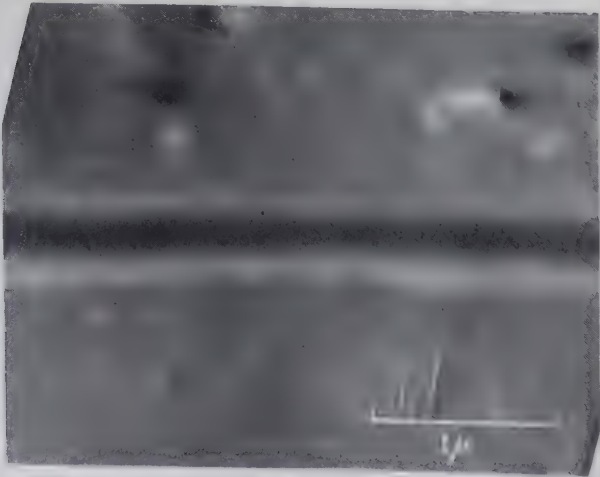
7a



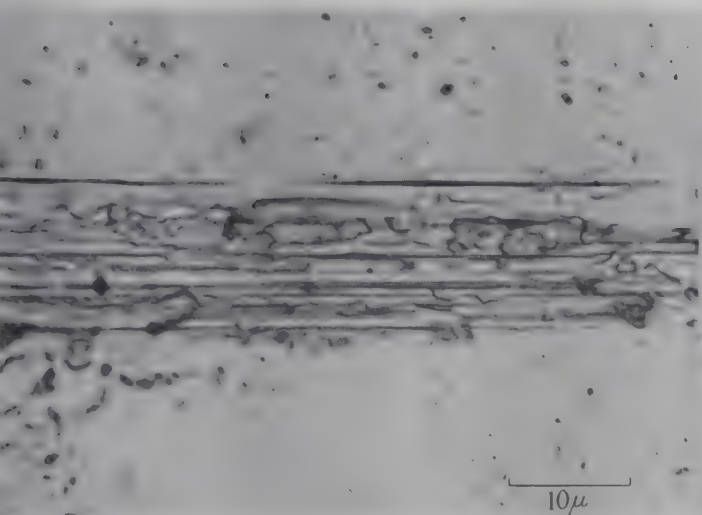
6b



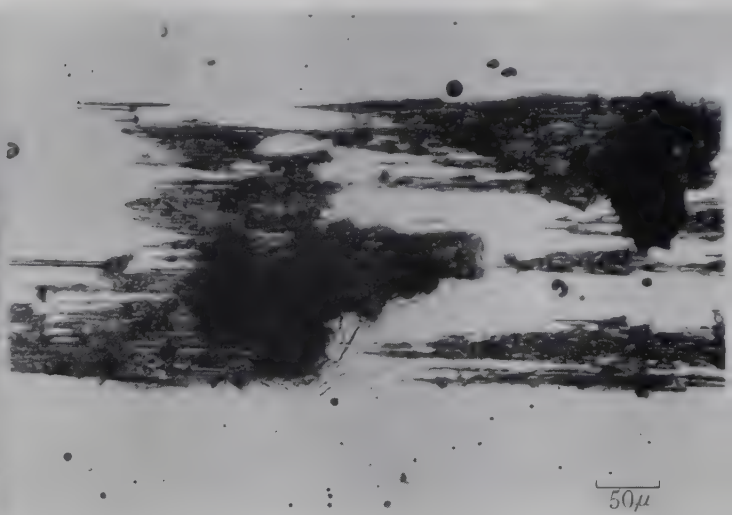
8b



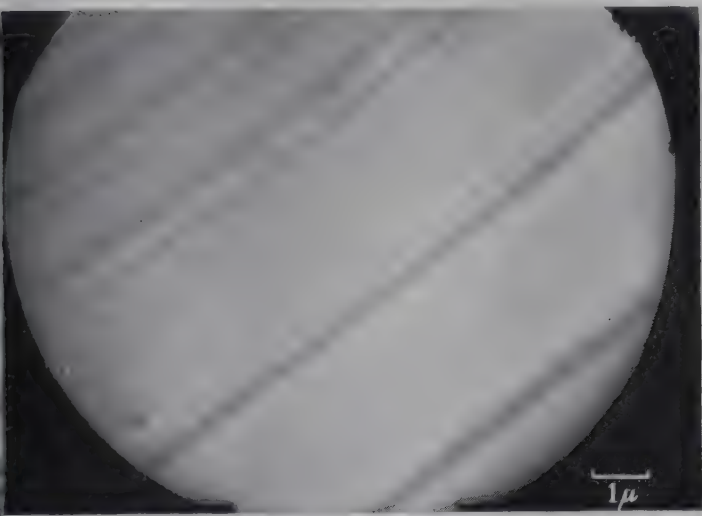
7



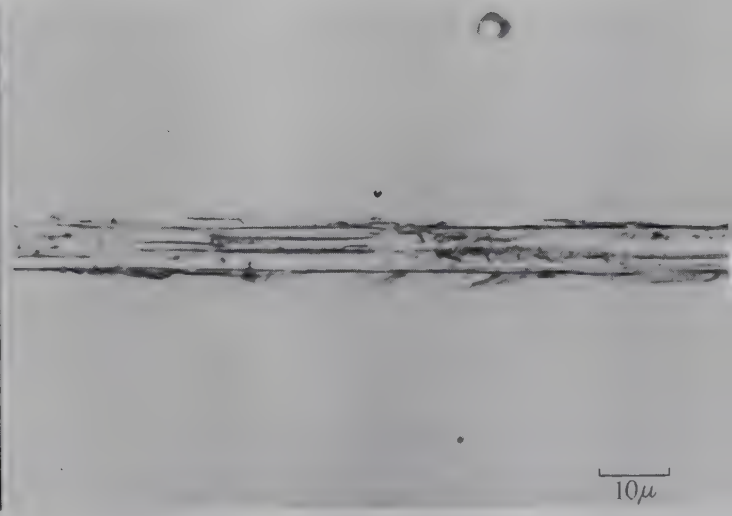
10a



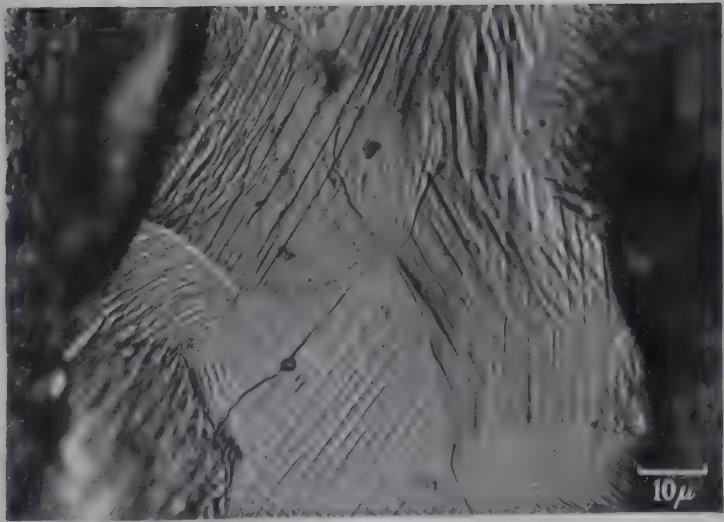
14a



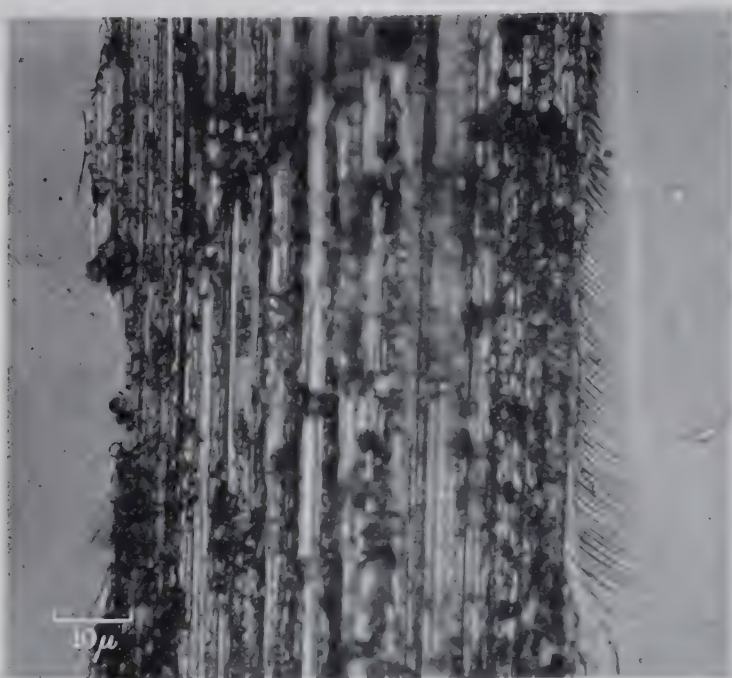
10b



14b



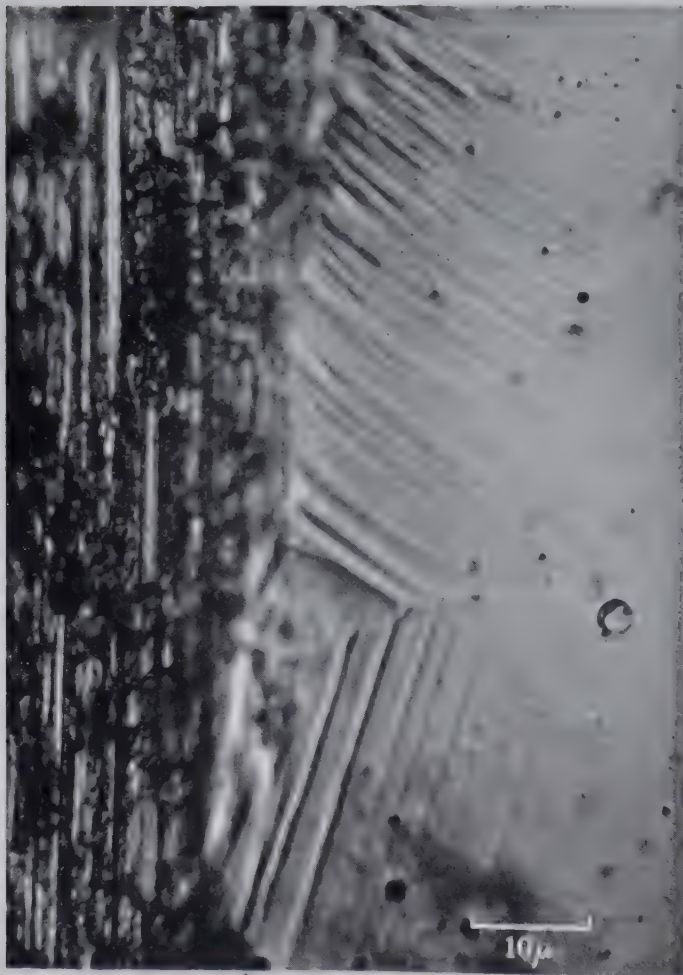
15



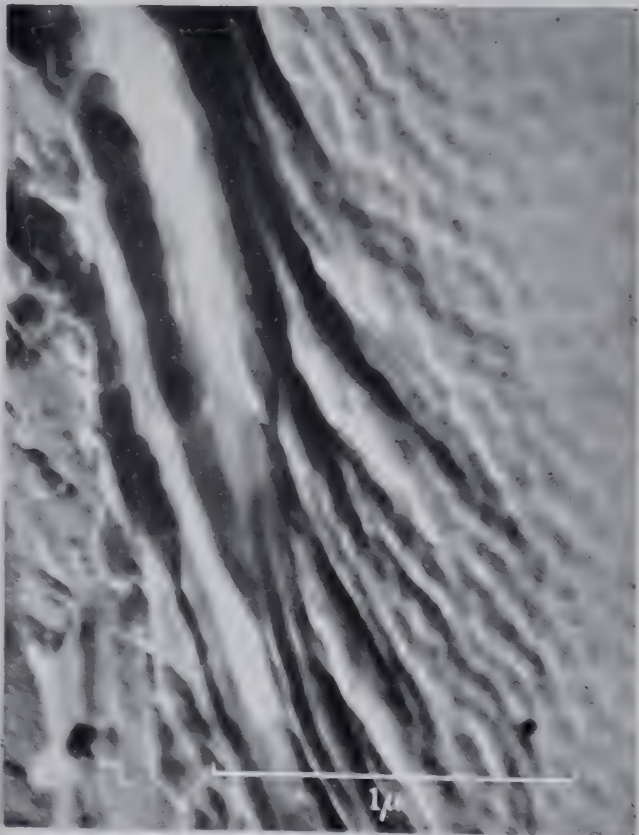
16a



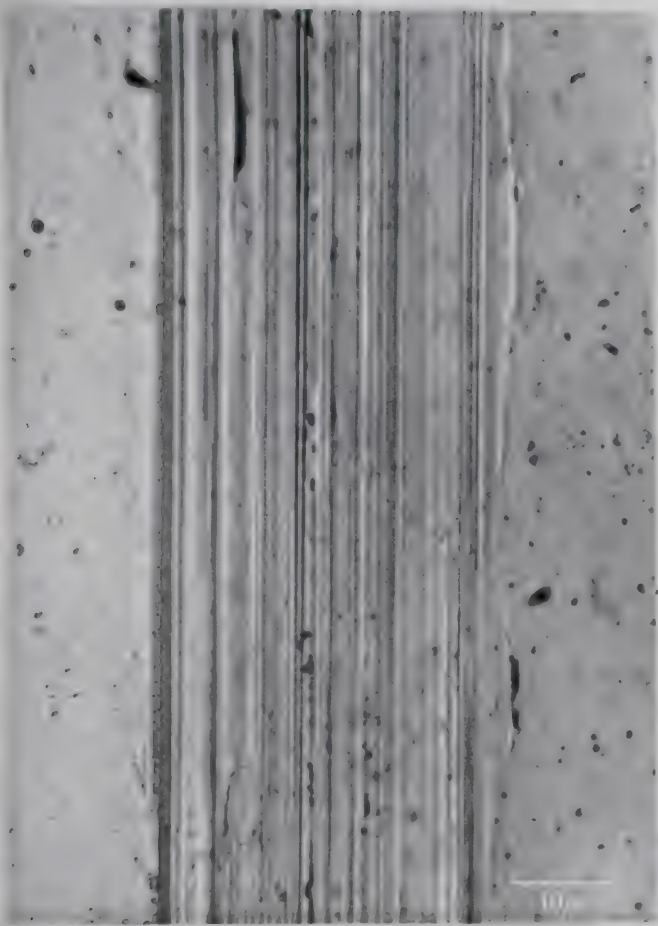
17a



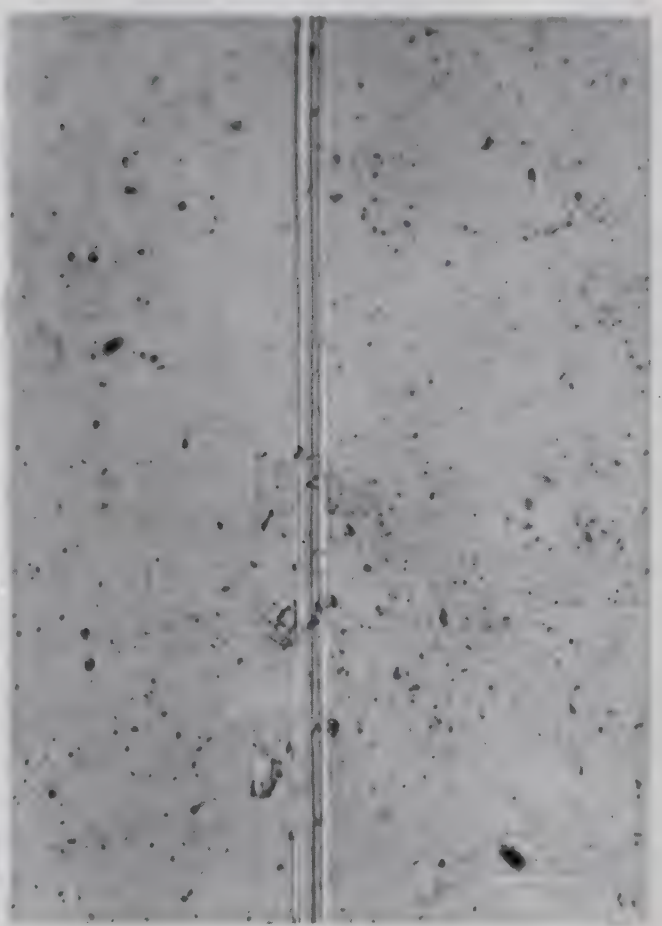
16b



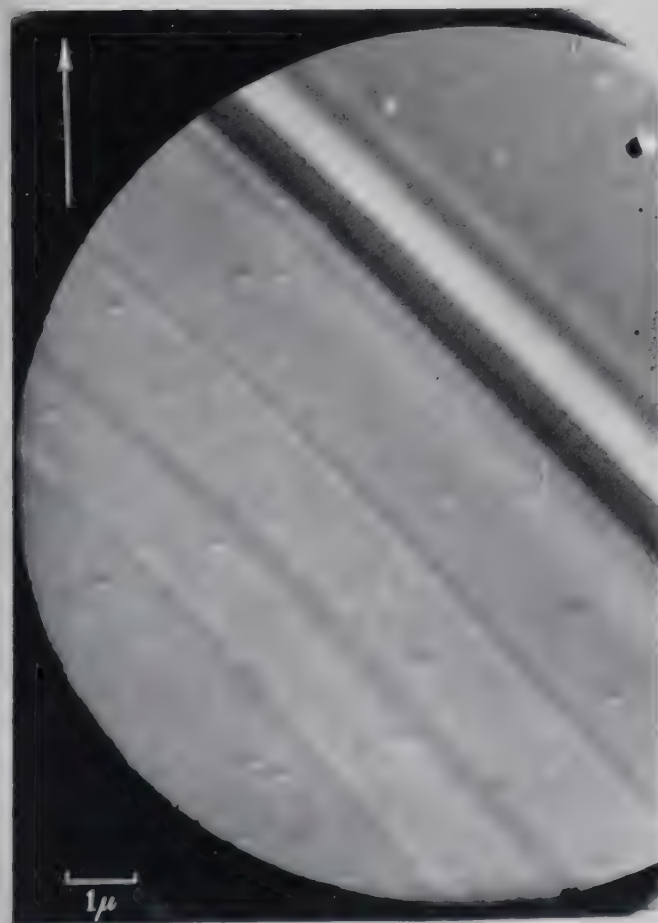
17b



19a



19b



19c



19d

For the purpose of examining the fine tracks of deformation due to sliding at very light loads, it was necessary to prepare a very much finer surface than that obtained by abrasion. Wherever possible, therefore, the surfaces were polished electrolytically using methods described elsewhere by Jacquet (1935, 1936) and later by Jacquet & Roquet (1939), Elmore (1939, 1940) and Honeycombe (1946). The technique differed slightly according to the metal used.

(3) *Examination of surface deformation*

The two principal methods used to examine the deformed surfaces were light and electron microscopy. The electron microscope was used to observe tracks in which the details were beyond the resolution limit of the light microscope.

Light microscopy involved no special technique. In order to observe the contour of metal surfaces in the electron microscope, however, it was necessary to produce thin film replicas which would permit the transmission of electrons. Several accepted replica techniques were used, notably that described by Mahl (1940) for aluminium. This technique involves the production and stripping of thin films of aluminium oxide. A method, perfected by Schaefer (1942), for producing thin robust films of polyvinylformal was used for metals other than aluminium and for non-metals.

Certain replicas were shadowed with a heavy metal in an attempt to increase the contrast and to provide a quantitative estimate of the size and nature of debris in the track. A mixture of gold and palladium was used in all cases. The presence of palladium tended to prevent the aggregation of gold atoms under the electron beam.

The effect of shadowing an aluminium oxide replica is shown in figure 4, plate 2. It represents a fine scratch on an aluminium surface. The pictures are electron micrographs of the same aluminium oxide replica before and after shadowing at 45° to the surface in the direction shown.

RESULTS

(1) *Copper*

The results of a large number of measurements of friction between a copper slider and a clean, dry electrolytically polished surface are summarized graphically in figure 5 (upper curve). The coefficient of friction is approximately constant (i.e. Amontons's law holds) between loads of 4 kg. and 40 g. Between 40 and 1 g. the friction falls to about one-quarter of its value at the higher loads. The range of loads within which the fall occurs is referred to hereafter as the region of transition. Amontons's law again appears to hold, at the low value of friction, for loads below 1 g.

The friction at loads within the region of transition was observed to fluctuate between the limits 0.4 and 1.8 defined by the upper and lower parts of the curve. The points on the curve in this region represent the mean of these values, but the appearance of the curve gives no indication of the large deviation from the mean.

The deformation of the surface at two values of load is shown in figure 6, plate 3. *a* is the track due to sliding at a load of 17 g. Damage to the surface is profound and the presence of slip bands, particularly at the right-hand edge of the track, provides evidence of the enormous local forces. It is clear that the metal has undergone

plastic flow due to the high local pressure. A calculation of that pressure, made on the assumption that contact has taken place over the whole track width (18μ) shows that it corresponds closely to the flow pressure of the metal.

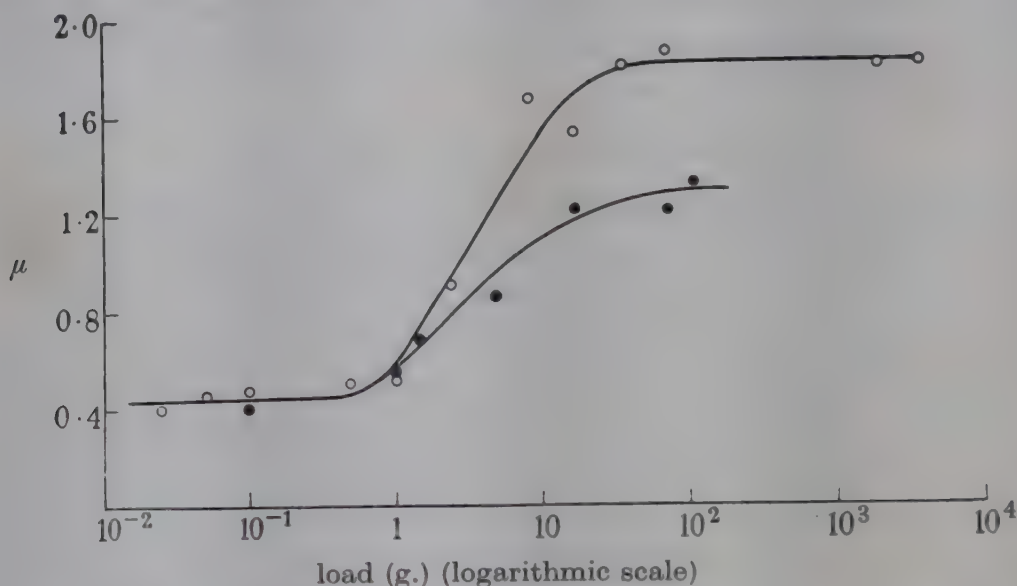


FIGURE 5. The friction of copper on electrolytically polished copper. ○, freshly prepared; ●, oxidized 400 to 600 Å.

The appearance of the deformation due to sliding at a load of 1 g., which is below the transition, is shown in figure 6*b*, plate 3. Grooves are apparent in the surface but there is little evidence of tearing. The tendency for tracks made at light loads to consist of several parallel grooves was noticed in all the experiments on copper. Its significance will be discussed.

Further evidence of the smooth nature of the track at light loads is provided by the electron micrograph of figure 7, plate 3. The micrograph was produced from a Formvar (polyvinylformal) replica of a track produced by sliding under a load of 0.4 g.

The picture indicates that the track consists of a smooth groove ploughed out of the surface with a ridge raised on either side. Smaller parallel grooves were observed which are not shown.

The reduction of friction at light loads was considered to be due to the thin (80 to 100 Å) layer of copper oxide which is present on the surface of copper in air at room temperature. The layer offered protection to the naked metal at light loads. The behaviour of the friction and the appearance of the corresponding deformation indicate that an increasing proportion of metallic junctions was formed at heavy loads, corresponding to progressive penetration of the oxide. The validity of Amontons's law for loads above the transition shows that the oxide film played little part in the process of sliding at heavy loads.

Thick oxide film. Further experiments were performed to observe the effect of increasing the thickness of the oxide film. An electrolytically polished copper specimen was oxidized to a thickness of 400 to 600 Å by heating in air. The thickness was estimated from the first order interference colours which have been tabulated by Constable (1929).

The results of the measurement of friction, at various loads, between a copper slider and the oxidized surface are given in the lower curve of figure 5. A transition from low to high values of friction occurred over approximately the same range of loads as before. The coefficient of friction at loads above the transition was reduced by oxidation whilst the friction below the transition was unchanged.

A corresponding change occurred in the deformation in the track of the slider. Figure 8*a*, plate 3, is an optical micrograph of a track caused by the passage of the slider under a load of 112 g. Figure 8*b*, plate 3, is an electron micrograph of the track produced on the same surface at a load of 0.15 g. Comparison with figure 6 shows that there was a reduction in the amount of deformation at loads below the transition on both types of surface.

Experiments with a steel slider. Similar experiments were performed using a steel slider on a clean, dry, electrolytically polished copper specimen. The results of the measurement of friction are given in figure 9.

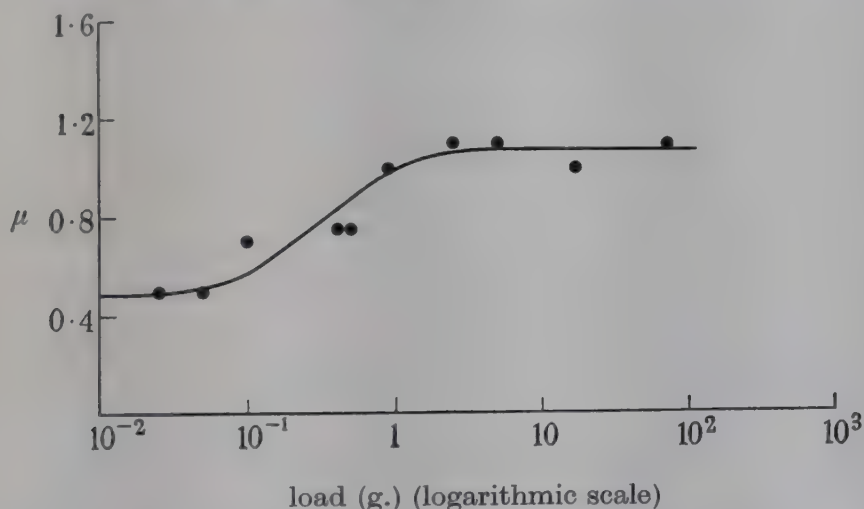


FIGURE 9. The friction of steel on electrolytically polished copper.

The friction of steel on copper at heavy loads was $\mu = 1.1$ compared with $\mu = 1.8$ for copper on copper. The friction was dependent on load, as before, and the coefficient of friction at loads below 0.1 g. was approximately equal to that for copper on copper. The transition to the low value was observed to occur between 2 and 0.1 g. The corresponding transition using a copper slider was between 20 and 1 g.

The deformation in the track at loads above the transition was less than that for copper on copper, as shown by the photomicrograph of figure 10*a*, plate 4. The electron micrograph of figure 10*b*, plate 4, shows that the deformation at a much lighter load was very similar to that caused by a copper slider (figure 7, plate 3). The corresponding value of friction was also the same.

For steel on copper, therefore, a transition in the value of friction and in the character of the surface deformation was still evident, though not so marked as it was for copper on copper.

The effect of surface roughness. The preceding experiments were made on copper surfaces which had been prepared by electrolytic polishing. The irregularities on such a surface are known to be very small and different in character from those on surfaces prepared in any other way. For instance, Raether (1947) has shown that

the surface asperities, if they may be so described, rise very gradually with a maximum gradient of approximately one in twenty. He suggests that there are unlikely to be steps on the surface of dimensions greater than a single atomic layer.

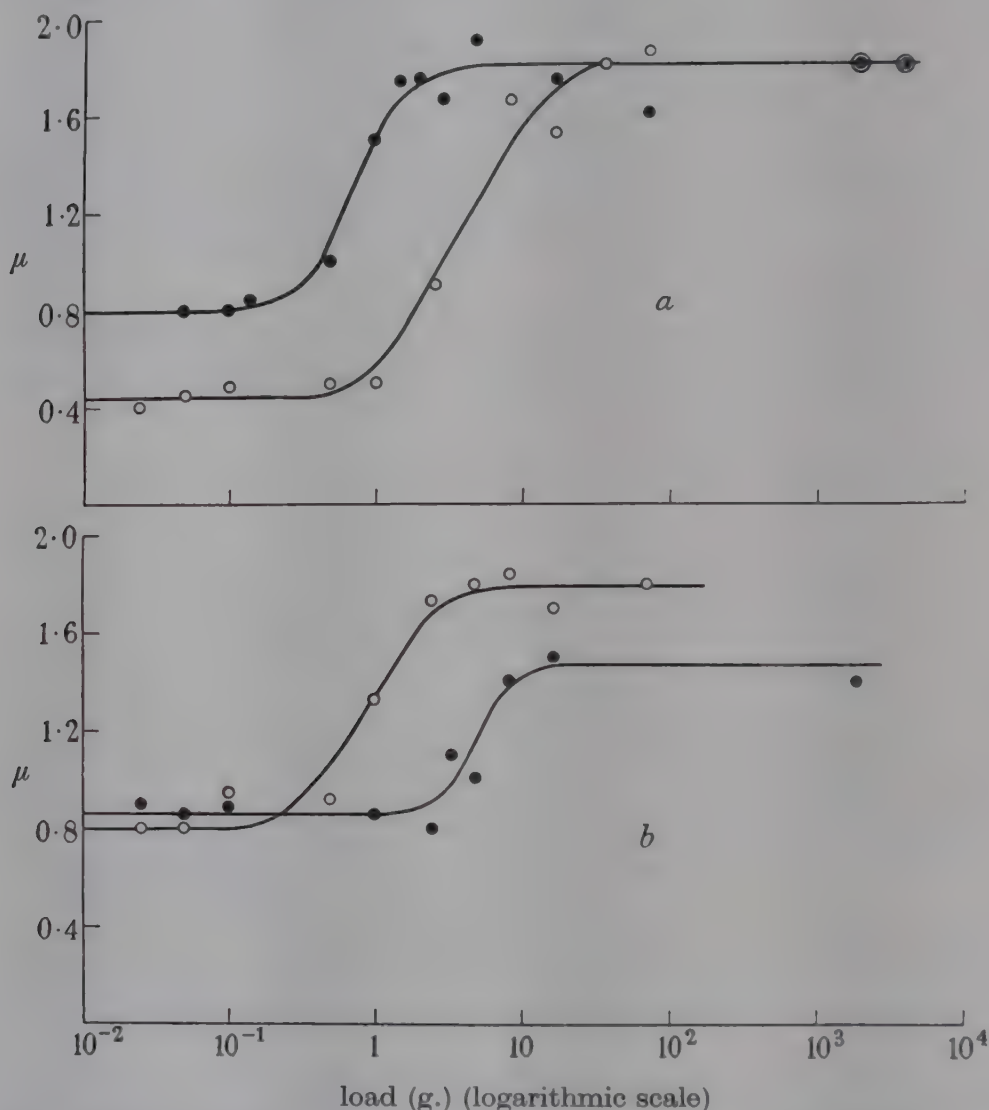


FIGURE 11. The influence of surface finish on the friction of copper on copper. *a*, electrolytically polished surfaces: $\circ$, smooth; $\bullet$, rough. *b*, mechanically prepared surfaces: $\circ$, mechanically polished; $\bullet$, abraded.

Experiments were carried out to compare the frictional behaviour of an electrolytically polished surface with that of surfaces of greater roughness. Alternative specimens were prepared by the following means. The estimated mean height of the surface asperities is given in brackets:

- (a) Abrasion under water on 600 grade emery paper (0.2μ).
- (b) Mechanical polishing on 'Selvyt' cloth impregnated with levigated alumina (0.05μ).
- (c) Deliberate pitting of an electrolytically polished surface. This was achieved by raising the current in the polishing cell until gas was just evolved at the surface of the specimen ($< 1.0\mu$).

Surfaces prepared by these methods were very much rougher than the good electrolytically polished surfaces previously described. Any influence of surface

roughness upon friction would, therefore, be expected to appear in the results of comparative measurements.

The results of friction measurements on the four types of copper surface are given in figure 11.

It is evident that the frictional process is very similar on all the specimens. Detailed differences in behaviour are apparent. In particular the lowest value of friction at heavy loads was found to occur on the abraded surface, whilst the electrolytically polished surface provided the lowest friction at loads below the transition. The implications of these results will be discussed.

(2) Silver

Measurements of friction between a silver slider and an abraded silver surface were made over the same range of loads as the measurements on copper. The results are presented in figure 12. It is evident that, within the limits of error, Amontons's law was obeyed down to the smallest load at which observations were made.

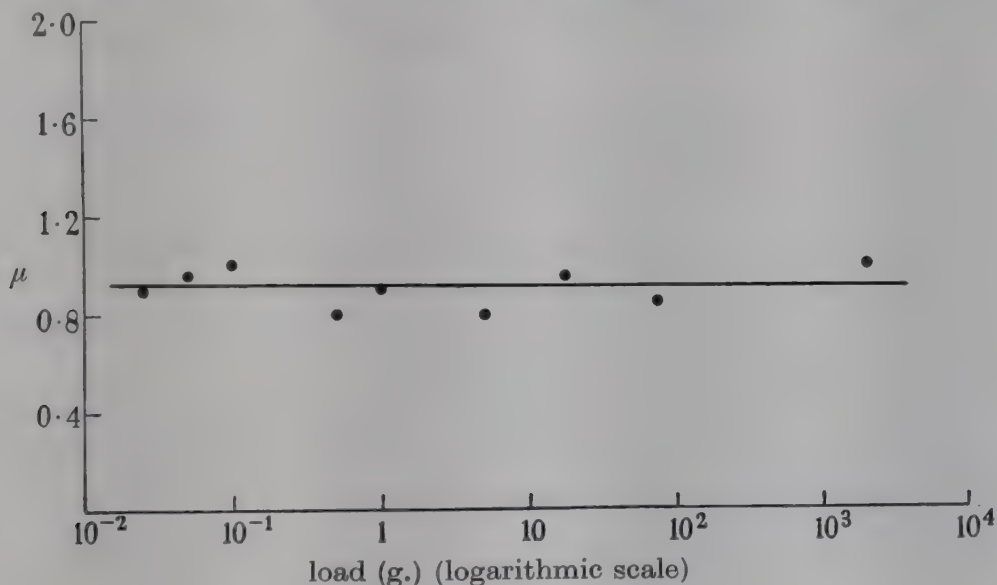


FIGURE 12. The friction of silver on silver.

(3) Aluminium

Further experiments were made by sliding an aluminium slider on an electrolytically polished aluminium surface. Figure 13*a* gives the results of the measurement of friction. Amontons's law was true ($\mu = \text{constant}$) for loads which varied by a factor of 10^5 . There was no tendency for the friction to be less at light loads, as it was for copper.

Figure 14, plate 4, shows the corresponding deformation on the track at two selected values of the load. The deformation is split into a number of tracks, each of which shows profound plucking of the aluminium and large particles are seen projecting from the surface. The intense local pressure has caused slip along the crystal planes of grains adjacent to the track. Figure 15, plate 4, shows this process more clearly. It is a photomicrograph of the area between two tracks caused by an aluminium slider at a load of 2 kg.

Slip lines are also just visible at the edge of the track caused by sliding at a load of 0.2 g. (figure 14*b*, plate 4). The damage within the track is still profound and

resembles the deformation of copper at much heavier loads (e.g. figure 6, plate 3). This agrees with the evidence that the friction is high at this value of load.

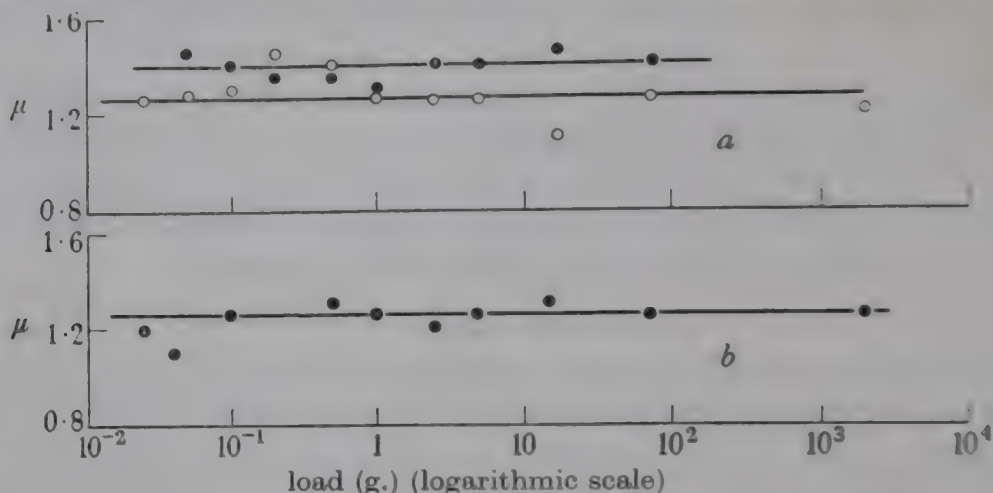


FIGURE 13. *a*, The friction of aluminium on aluminium. *b*, the friction of steel on aluminium: ○, freshly prepared; ●, oxidized 750 Å.

The friction and deformation due to a steel slider on aluminium are very similar. The friction is again the same at all values of load (figure 13*b*). Optical micrographs of the deformation at two high values of load (160 g., 20 g.) are given in figure 16, plate 5. The slip lines appear particularly clearly. The deformation at a light load (0.4 g.), recorded in the electron micrographs of figure 17, plate 5, is strikingly similar. This shows that the mechanism of sliding on aluminium is the same at all loads within the range of the present experiments.

Thick oxide film. Measurements of friction on electrolytically polished aluminium which had been oxidized anodically to a depth of about 1000 Å are also presented in figure 13*a*. There was very little change in the value of friction and Amontons's law was observed to hold, as before.

(4) *Friction and deformation of non-metals*

A few experiments were performed to measure the friction at light loads on two non-metals, namely diamond and synthetic sapphire. The slider used was of the same material as the lower surface in each experiment. The results for diamond showed that Amontons's law held approximately over the range of loads used in the experiments, from 0.1 to 100 g., with a coefficient of friction about 0.12. No damage to the diamond surfaces was observed.

The friction of sapphire on sapphire was higher ($\mu = 0.3$) and there was slight surface deformation.

Although only a few experiments were made the results are given here because of the practical interest in the frictional properties of non-metals for the design of instrument bearings.

(5) *Lubrication by long-chain polar compounds*

A restricted series of experiments on electrolytically polished copper lubricated by long-chain fatty acids showed a marked deviation from Amontons's law at loads below 1 g.

Only two fatty acids were used, namely lauric acid ($C_{11}H_{23}COOH$) and octacosanoic acid ($C_{27}H_{55}COOH$). In the first and most extensive series of experiments each was applied to the surface as a monolayer by retraction from a molten crystal in the manner described by Bigelow, Glass & Zisman (1947).

Although the surface treated in this way was indistinguishable in appearance from the clean surface the friction was reduced by a factor of more than 10. The deformation of the surface was recorded in the usual way by photo- and electron micrography.

The results of measurements of friction (figure 18) show a definite trend towards higher values of friction at light loads. The trend is more marked in the longer chain compound, octacosanoic acid (dotted curve). The sliding at a load of 17 g. was very smooth and the coefficient of friction remained fairly constant between 0.16 and 0.20. At a load of 0.025 g., however, wild fluctuations ('stick-slip') occurred and the value of the coefficient measured at the peaks varied from 0.5 to 1.1. The dotted curve in figure 18 is drawn through the mean of these peaks and the vertical lines represent the extent of the fluctuation.

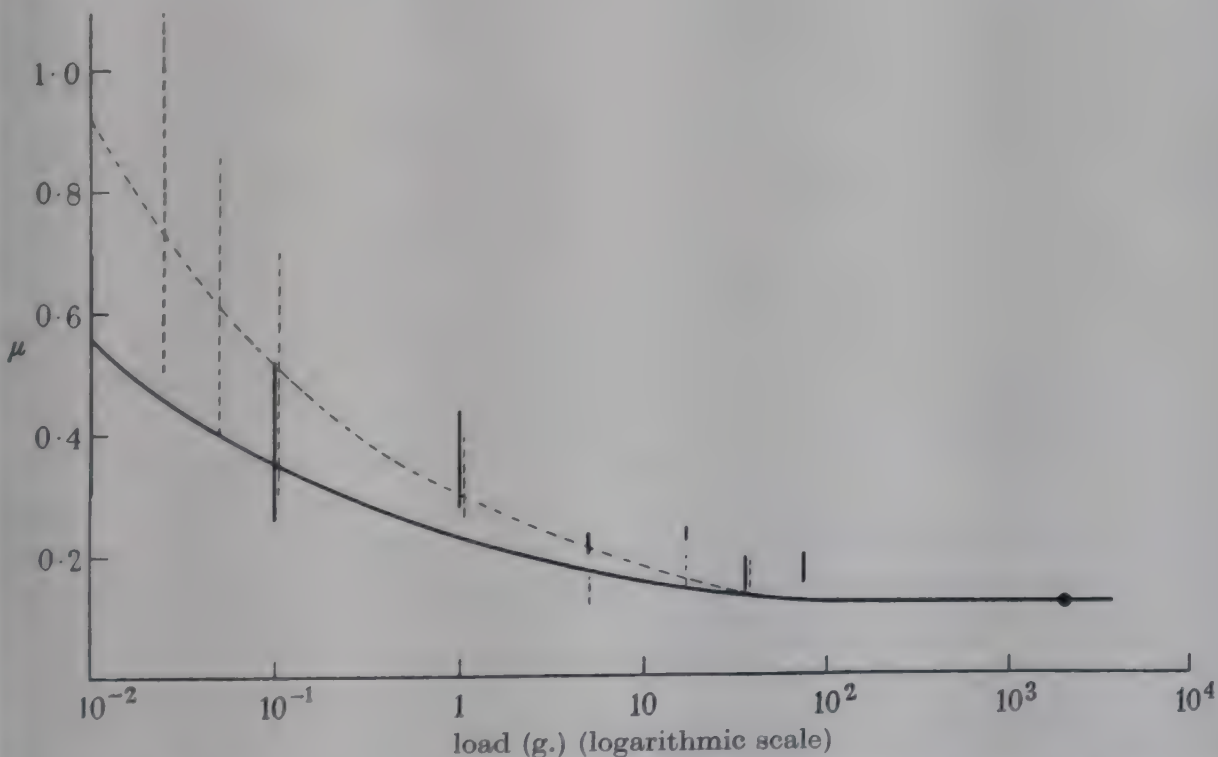


FIGURE 18. The friction of copper on copper boundary lubricated by a fatty acid. The vertical lines represent the extent of the fluctuations in each measurement. —, lauric acid;, octacosanoic acid.

The order of sliding was varied in subsequent experiments in order to eliminate the effect of ageing of the monolayer. Similar results were obtained.

Figure 19, plate 6, shows that the deformation of the surface was similar at all values of load. *a* and *b* are photomicrographs corresponding to normal loads of 17 and 0.5 g. respectively. *c* and *d* are electron micrographs of 'Formvar' replicas taken from a surface deformed under the latter load (0.5 g.). All the micrographs relate to lubrication by lauric acid.

DISCUSSION

(1) *The influence of oxide films*

The experimental results show that the friction of a copper or steel slider on a copper surface depends upon the normal load applied to the sliding surfaces. Amontons's law holds at very high and very low loads, but with different values of the coefficient of friction. These asymptotic values depend somewhat on the method of preparation of the surface.

The transition from one value of friction to the other occurs within a range of loads from 1 to 10 g. The character of the deformation in the track of the slider shows a corresponding change, which occurs abruptly at a load in the region of altering friction. It is therefore reasonable to attribute the two changes to a single cause.

The appearance of the surface after sliding at loads above the transition leaves little doubt that intimate metallic contact has occurred and that the friction corresponds to the sliding of clean surfaces in air. The appearance of the surfaces after sliding at loads below the transitional value gives the impression of lubricated sliding. The tracks are smooth (figures 6 and 7, plate 3) and there is no sign of debris or of broken welds. The fact that the coefficient of friction is lower in the latter case is confirmation of the theory that the frictional force between two metals is largely the force required to shear welded junctions which are formed during sliding.

It is evident that a surface film is intervening to protect the naked metal at the lighter loads. The low value of friction corresponds to sliding on this film, not on the metal. The transition corresponds to penetration of the film which is, naturally, spasmodic at the critical load. This accounts for the fluctuating values of friction in the region of transition. The high value of friction at heavier loads corresponds to true metallic sliding in which the film plays little part.

It is known that a layer composed mainly of oxide (Cu_2O) and probably some sulphide (Cu_2S) and other compounds is formed when a freshly-prepared copper surface is exposed to the atmosphere of the laboratory. Miley & Evans (1937) have estimated that the thickness of such a film is approximately 80 Å on an abraded copper surface. Pilling & Bedworth (1923) have pointed out that the oxide film is formed under compression, due to the expansion in volume which occurs on the formation of the oxide. Electron diffraction experiments by Finch & Quarrel (1933) have confirmed this view. Further confirmation is provided by Evans (1946) who has observed that such films wrinkle when they are removed from the metal surfaces.

It is apparent that the nature of the first few molecular layers above the metal surface might also depend upon the method of preparation. Although auxiliary films of this kind may be present, the striking similarity of the friction-load curves for surfaces prepared in different ways indicates that the film which is produced by exposure to the atmosphere is responsible for the main effect.

The experimental results may, therefore, be explained in terms of the intervention of the surface film into the frictional process. At light loads the thin film prevents inter-metallic contact. Experiment shows that it begins to break down at a load of about 1 g. on an electrolytically polished surface. Metallic junctions then begin to

be formed. The exact value of the load at which penetration first occurs is connected with the amount of deformation caused by plastic flow under the high local pressure. This may well be influenced by the distribution of the points of contact which, in turn, depends upon the macroscopic contour of the sliding surfaces. This provides an explanation of the observed variations in the load at which transition occurs on different types of surface and in different experiments.

The reason for the proportionality of friction to load, with two different values of the coefficient of friction, is readily seen. The real area of contact at any load is determined by the plastic deformation of the surfaces. It is determined by the load and is proportional to it. The frictional force is required to shear this area of material. The friction is constant provided the shearing always takes place in the same material, be it metal or oxide. The former corresponds to high values of friction, the latter to low values. At intermediate loads the slider breaks through the surface film only sporadically and the friction consequently fluctuates. The departure from Amontons's law does not arise from a change in the fundamental mechanism of friction but is due to an increase in the relative importance of the thin film at light loads.

These conclusions concern only a clean copper surface covered by the thin coherent film which is formed by exposure to the atmosphere at room temperature. We may now consider the effect of a thick layer of oxide formed by deliberate oxidation at a high temperature. The results show that a layer of oxide 400 to 600 Å thick reduces the friction at heavy loads from $\mu = 1.8$ to $\mu = 1.3$ whilst leaving the friction at light loads the same ($\mu = 0.4$). The fact that penetration of the oxide layer occurs at about the same load on either the thick or the thin layer suggests that the additional layer of oxide behaves in a different way from the thin coherent film formed at room temperature. The thin film is still primarily responsible for protection or resistance to penetration. The thick layer only reduces the friction by virtue of its bulk. The relatively large quantity of oxide in the thicker film tends to form inclusions which weaken the intermetallic bond, thus reducing the friction at heavy loads.

The method of preparation of the surface was observed to have some influence upon the friction. It is probable that the thickness of the oxide layer depends upon the method by which the surface is prepared. Moreover, the volume of oxide contained within a given apparent area of the surface clearly depends upon the roughness of the surface which governs the real area exposed to the atmosphere. The volume is greater for an abraded specimen than for one that is electrolytically polished. An abraded surface is also more favourable to the formation of inclusions. These considerations might explain why the friction at heavy loads is less on an abraded (rough) surface than on an electrolytically polished (smooth) one, as in figure 11.

It must be emphasized that this discussion relates to sliding in air. Other workers have shown that the friction between degassed metals *in vacuo* can be very much greater than the values recorded here. For instance, Bowden & Hughes (1939) measured a value of $\mu = 4.8$ for copper on copper under clean conditions *in vacuo*. This fell to $\mu = 2.4$ on admission of a trace of oxygen. Tingle (1950) endeavoured to remove the oxide layer by cutting the surface immediately in advance of the slider

but found no appreciable reduction of friction when exposure to the atmosphere was reduced to four seconds. Sliding is, however, a dynamic process. Deformation of the surface occurs some little way ahead of the slider and the oxide film is therefore penetrated and broken up. The naked metal thus revealed comes into contact with the atmosphere before the slider reaches it and there is time for a layer of oxygen to be adsorbed. The slider will be unable to reach a part of the surface that is free from contamination by atmospheric gases, so long as sliding takes place in air. This is the reason why friction between metals in air remains low compared with that between metals degassed *in vacuo*.

It is consequently unlikely that cutting the surface in advance of the slider will remove or reduce the effect of an adsorbed film.

In contrast to the behaviour of copper, the experimental results on silver and aluminium show that the mechanism of sliding on these metals is the same over the whole range of loads from 0.01 g. to 4 kg. The departure from Amontons's law for copper was ascribed to the effect of a surface film of oxide in separating the metal surfaces at light loads. The copper oxide was, in fact, acting as a boundary lubricant.

The behaviour of silver indicates that there is no coherent film of oxide capable of protecting the naked metal. This is consistent with the evidence (e.g. Evans 1946) that the oxide on a freshly prepared specimen of silver is very thin. Aluminium, however, is known to have a coherent film of oxide of comparable thickness to that on copper. The different behaviour of these two metals must be attributed to the contrasting properties of their oxides.

In order to produce a low value of friction a surface film should resist penetration and at the same time be easily sheared. Copper oxide appears to fulfil both these requirements. Aluminium oxide must fail in one or both respects.

Aluminium oxide (Al_2O_3) is very hard whilst the hardness of copper oxide is comparable with that of the metal, as shown in the table below:

metal	hardness (Moh's scale)	oxide	hardness (Moh's scale)
copper	3.0	cuprous oxide	3.5 to 4.0
aluminium	2.0 to 2.5	aluminium oxide	9.0

On copper, therefore, it is possible that for small amounts of deformation the metal and oxide undergo plastic deformation together. There is no such possibility with aluminium. In all cases the metal will deform below the film long before the pressure is sufficient to produce plasticity in the oxide. This will tend to undermine the oxide film and may well be the determining factor in the process of disruption.

Another factor which may influence the protective behaviour of an oxide film is the degree of fit or misfit which exists between the crystal lattices of oxide and metal and the security of attachment of the oxide to the metal.

The thin film of oxide on copper is formed under compression because the crystal lattice spacing of the oxide is about 18% greater than that of the metal. The oxide is crystalline (Darbyshire 1931) and is probably oriented with the grains of the underlying metal. The stability of thin crystalline films formed under conditions of misfit compression is discussed by van der Merwe (1949), and by Cabrera & Mott (1949).

On the other hand there is no evidence of orientation of the oxide on aluminium (Bircumshaw & Preston 1936). The oxide appears to grow as an independent but coherent film and may only be loosely attached to the surface of the metal. These facts, combined with the difference in the bulk properties of the oxides of aluminium and copper may well account for the contrasting frictional properties of the two metals at light loads. Further knowledge about the nature of thin oxide films is required before the mechanism of sliding can be explained in more detail.

(2) *The influence of lubricant films*

The preliminary experimental results of sliding copper on copper lubricated by a monolayer of a long-chain polar compound, show a departure from Amontons's law such that the friction becomes greater at the lighter loads. The effect appears to increase with the length of the molecular chain for the two compounds studied (lauric and octacosanoic acids).

The departure from Amontons's law suggests the presence of a small force additional to the 'true' frictional force. This may be a measure of the forces of adhesion between the surface films of lubricant. An estimate of the force of adhesion, if such it is, would be important because of the information it would provide about the properties of these monomolecular layers. Although it is not possible to make such an estimate, because of the uncertainty of the actual area of contact over which adhesion occurs, it is nevertheless clear that the coefficient of friction is greater at very light loads. This fact is of great practical interest because of its relevance to the design of very small lubricated bearings.

I thank Dr F. P. Bowden, F.R.S., for advice and encouragement in the work; Dr D. Tabor for helpful discussions; the Ministry of Supply (Air) for grants to the Laboratory; the Telecommunications Research Establishment from which I am seconded; the Cavendish Laboratory and Dr V. E. Cosslett for use of the electron microscope; Dr A. F. Brown for preparing aluminium oxide replicas; finally, the Chief Scientist, Ministry of Supply, for permission to publish this paper.

REFERENCES

- Bigelow, W. C., Glass, E. & Zisman, W. A. 1947 *J. Coll. Sci.* **2**, 563.
 Bircumshaw, L. L. & Preston, G. D. 1936 *Phil. Mag.* **22**, 654.
 Bowden, F. P. & Hughes, T. P. 1939 *Proc. Roy. Soc. A*, **172**, 265.
 Bowden, F. P. & Leben, L. 1939 *Proc. Roy. Soc. A*, **189**, 371.
 Bowden, F. P. & Tabor, D. 1939 *Proc. Roy. Soc. A*, **169**, 391.
 Bowden, F. P. & Tabor, D. 1943 *J. Appl. Phys.* **14**, 141.
 Bowden, F. P. & Tabor, D. 1946 *Ann. Rep. Chem. Soc.* **42**, 20.
 Cabrera, N. & Mott, N. F. 1949 *Rep. Progr. Phys.* **12**, 163.
 Constable, F. H. 1929 *Proc. Roy. Soc. A*, **125**, 630.
 Darbyshire, J. A. 1931 *Trans. Faraday Soc.* **27**, 675.
 Elmore, W. C. 1939 *J. Appl. Phys.* **10**, 724.
 Elmore, W. C. 1940 *J. Appl. Phys.* **11**, 797.
 Evans, U. R. 1946 *Metallic corrosion, passivity and protection*. London: Arnold.
 Finch, G. I. & Quarrel, A. G. 1933 *Proc. Roy. Soc. A*, **141**, 398.
 Honeycombe, R. W. K. 1946 *C.S.I.R. Phys. Metallurgy Report*, No. 1 (A 146).
 Holm, R. 1946 *Electric contacts*. Hugo Gebers Förlach.
 Jacquet, P. 1935 *C.R. Acad. Sci., Paris*, **201**, 1473.

- Jacquet, P. 1936 *Bull. Soc. chim. Fr.* **3**, 705.
 Jacquet, P. & Roquet, P. 1939 *C.R. Acad. Sci., Paris*, **208**, 1012.
 Mahl, H. 1940 *Metallwirtschaft*, **19**, 488.
 van der Merwe, J. H. 1949 *Discussion Faraday Soc.* No. 5, p. 201.
 Miley, H. A. & Evans, U. R. 1937 *J. Chem. Soc.* p. 1295.
 Moore, A. J. W. 1948 *Proc. Roy. Soc. A*, **195**, 231.
 Pilling, N. B. & Bedworth, R. E. 1923 *J. Inst. Metals*, **29**, 529.
 Raether, H. 1947 *Metaux et Corrosion*, **22**, 257, 2.
 Schaefer, V. J. 1942 *Phys. Rev.* **62**, 495.
 Tingle, E. D. 1950 *Trans. Faraday Soc.* **46**, 93, 97.

DESCRIPTION OF PLATES 2 TO 6

Note. The arrows on electron micrographs indicate the direction of shadowing with a heavy metal.

PLATE 2

FIGURE 2. View of friction apparatus showing arrangement of the pick-up head relative to the slider and specimen. The micrometer in the foreground is used for calibration in the friction plane.

FIGURE 4. Electron micrographs ($\times 16,000$) of an aluminium oxide replica. *a*, before and *b*, after shadowing with gold and palladium at 45° to the surface in the direction shown. The picture represents a track made by a steel slider on aluminium under a load of 0.4 g.

PLATE 3

FIGURE 6. Optical micrographs ($\times 1150$) of deformation to electrolytically polished copper due to the passage of a copper slider at the following loads: *a*, 17 g.; $\mu = 1.5$. *b*, 1 g.; $\mu = 0.5$.

FIGURE 7. Electron micrograph ($\times 20,000$) of the deformation to electrolytically polished copper due to the passage of a steel slider at a load of 0.5 g.

FIGURE 8. Tracks caused by a copper slider on copper surface which has been oxidized by heating in air. *a*, optical micrograph ($\times 330$); load 112 g.; $\mu = 1.3$. *b*, electron micrograph of Formvar replica ($\times 2250$); load 0.5 g.; $\mu = 0.5$.

PLATE 4

FIGURE 10. Deformation due to a steel slider on electrolytically polished copper. *a*, optical micrograph ($\times 1250$); load 2.5 g.; $\mu = 1.1$. *b*, electron micrograph of Formvar replica ($\times 6000$); load 0.4 g.; $\mu = 0.75$.

FIGURE 14. Optical micrographs of deformation to electrolytically polished aluminium caused by an aluminium slider. *a* ($\times 140$); load 74 g.; $\mu = 1.3$. *b* ($\times 750$); load 0.2 g.; $\mu = 1.3$.

FIGURE 15. Optical micrograph ($\times 700$) showing the deformation of an electrolytically polished aluminium surface between two tracks, each due to the passage of an aluminium slider under a load of 2 kg. Slip along the grain boundaries and within the grain is evident.

PLATE 5

FIGURE 16. Optical micrographs of tracks caused by a steel slider on electrolytically polished aluminium. *a* ($\times 800$); load 20 g.; $\mu = 1.3$. *b* ($\times 1200$); load 100 g.; $\mu = 1.3$.

FIGURE 17. Deformation due to a steel slider on electrolytically polished aluminium. Electron micrographs of oxide replicas shadowed at 45° to the surface with gold and palladium. *a* ($\times 10,000$); load 0.4 g.; $\mu = 1.3$. *b* ($\times 37,500$); load 0.4 g. (edge of track); $\mu = 1.3$.

PLATE 6

FIGURE 19. Tracks due to copper sliding on electrolytically polished copper, lubricated by a monolayer of lauric acid. *a*, optical micrograph ($\times 1050$); load 17 g. *b*, optical micrograph ($\times 1050$); load 0.5 g. *c*, electron micrograph of Formvar replica ($\times 6000$); load 0.4 g. *d*, electron micrograph of Formvar replica ($\times 16,500$); load 0.4 g.

Electronic wave functions

II. A calculation for the ground state of the beryllium atom

By S. F. BOYS, *Theoretical Chemistry Department, University of Cambridge**

(Communicated by Sir Alfred Egerton, F.R.S.—Received 31 August 1949)

An approximate wave function expressed in terms of exponential functions, spherical harmonics, etc., with numerical coefficients has been calculated for the ground state of the beryllium atom. Judged by the energy criterion this gives a more accurate result than the Hartree result which was the best previously known. This has been calculated as a trial of a fresh method of calculating atomic wave functions. A linear combination of Slater determinants is treated by the variational method. The results suggest that this will provide a more powerful and convenient method than has previously been available for atoms with more than two electrons.

1. INTRODUCTION

A numerical calculation of an approximate wave function for the ground state of the beryllium atom will be described here. The method used consisted of taking a linear combination of up to ten Slater determinants and adjusting the linear coefficients by the ordinary variational procedure. The orbitals of the Slater determinants were products of simple radial exponential functions, spherical harmonics, and spin functions. If the quality of such wave functions is judged by the generally accepted test of the lowness of the calculated energy, the function found by this procedure is significantly better than all previous functions for this atom and state.

Such a general method has not been used previously, although the idea of using more than one Slater determinant has been variously discussed, generally under the title of 'superposition of configurations'. The treatment of various states of the oxygen atom by D. R. Hartree, W. Hartree & Swirles (1939) corresponds most nearly to the present method. In this case effectively two Slater determinant terms were used and the orbitals were constructed from purely numerical functions to obtain the lowest calculated energy, using a modification of well-known iterative integration of the differential equations developed by Hartree. The results of the present investigation suggest that by the use of larger numbers of determinants with orbitals of exponentials, etc., less numerical work is necessary for the same degree of accuracy, and it is probably considerably easier to extend to higher degrees of accuracy. In addition, atomic properties can be calculated from these functions without numerical integrations.

This calculation was made not only to develop a method of calculating general atomic wave functions, but also to gain experience and improve detailed techniques in some aspects of the general methods discussed in part I (Boys 1950) for molecules as well as atoms. The general theory of the method is just the same as that of the most general method for molecules discussed there, except that exponential functions around a single centre are used instead of the general Gaussian functions. The numerical calculation of the necessary integrals is much easier in this case than in the

* Formerly I.C.I. Research Fellow, Imperial College, London.

molecular case, and it appeared most practical to improve the efficiency of the detailed techniques in the parts of the calculation common to both cases as far as possible before making any extensive calculations for the molecular case.

The general method was to commence with a complete class of three-dimensional functions. In the present case we take $r^{p+l}e^{-\alpha r}S^{lm}(\theta, \phi)$, where p is any integer, α is any positive real number, and $S^{lm}(\theta, \phi)$ is any spherical harmonic of degree l . A finite number n of these are taken and n orthogonal linear combinations of these constructed. All possible Slater determinants are constructed from these together with the single-electron spin functions, and the best linear combination of these determinants to represent the wave function is then found by the Ritz variational method. To carry out the complete converging process we should test different sets of n initial functions and select the best, or at least one which is better than that of the first n terms of an ordered complete system of functions. Then if this process is repeated for larger values of n we have a convergent process. In practice we have to work to some arbitrary standard of accuracy, and for the present we shall be satisfied if we surpass by a significant amount the Hartree type functions which have hitherto been the most accurate available.

It is well known that for spherically symmetrical systems such as atoms the complexity of the problem can be considerably reduced by taking the functions for the variational problem to be eigenfunctions of the operators L^2 , L_z , S^2 , and S_z , where L and S are used to represent the total orbital and spin angular momentum respectively. Instead of taking the Slater determinants in the present case we can take an exactly equivalent set of linear combinations of them which are simultaneous eigenfunctions of these operators. The actual stationary states of the atoms can be regarded as corresponding to definite sets of the eigenvalues, and it can be shown that to approximate to any selected stationary state we need only use the linear combinations of the Slater determinants which have exactly this particular set of eigenvalues. In this particular calculation it is known that the eigenvalues of the ground state of the beryllium atom are (0, 0, 0, 0), so that we shall only use the corresponding linear combinations. The technique of constructing these linear combinations is well known; it is known as the process of forming vector coupled functions (see Condon & Shortley 1935). In this programme we shall be repeatedly using and discussing these functions, and it appears well worth while for the sake of brevity and conciseness to introduce a synthetic word to describe these. We propose to use 'co-detors', an extension of the detors of part I, to describe a linear combination of Slater determinants, formed from orthogonal single-electron functions, which is a simultaneous eigenfunction of the operators L^2 , L_z , S^2 and S_z .

All other aspects of the method are exactly the same as are generally used for such wave functions. We shall use the ordinary many-electron Schrödinger Hamiltonian in atomic units given by

$$\begin{aligned} H &= -\frac{1}{2} \sum_i \nabla_i^2 - Z \sum_i \frac{1}{r_i} + \frac{1}{2} \sum_{i,j} \frac{1}{r_{ij}} \\ &= \sum_i K_i - Z \sum_i V_i + \frac{1}{2} \sum_{i,j} M_{ij}, \end{aligned} \quad (1)$$

where it is convenient for subsequent work to define K , V and M as shown. Relativistic and spin terms are neglected although spin is fully included in the wave function, since this is a function of x_i, y_i, z_i , and v_i ($i = 1$ to N the number of electrons),

where v_i are variables which can only take the two values $\frac{1}{2}$ and $-\frac{1}{2}$. We shall use $(f|a|g)$ and $(f|g)$ to denote the results of taking f^*ag and f^*g respectively, and integrating over all continuous variables and summing over all discrete spin variables, whatever the numbers of variables involved. We shall use Φ_r to denote the co-detors used as functions for the final variational problem, which consists of the solution of

$$\sum_s H_{rs} Y_s - E Y_r = 0, \quad (2)$$

where $H_{rs} \equiv (\Phi_r | H | \Phi_s)$, to give $\sum_s Y_s \Phi_s$ and E as the final wave function and energy.

In § 2 will be described the theoretical evaluation of the H_{rs} in terms of integrals between the orthogonal single electron functions. In § 3 the algebraic formulae for the various necessary integrals between the elementary single-electron functions before orthogonalization will be obtained. In § 4 the preparation of the orthogonal linear combinations of the single-electron functions and the numerical evaluation of the integrals between these will be discussed. In § 5 the numerical solution of the eigenvector problem will be described. In § 6 various numerical data of the calculation will be reported, and in § 7 the results will be discussed.

2. THE EVALUATION OF THE MATRIX ELEMENTS FOR THE VARIATION PROBLEM

It is necessary to define the co-detors which were used in this calculation and then to evaluate the integrals of the Schrödinger-Hamiltonian between them. These co-detors were constructed from a set of normalized orthogonal single-electron functions, that is, functions of x, y, z and v , which we shall represent by sA, sB, sC and pA . The sA is to be regarded as a single symbol, not a product of s and A , and similarly for the other symbols. The s and p denote that the angular factors of the functions are respectively spherical harmonics of degree 0 and 1. Each symbol represents a set of functions which have their dependence on r given by the same factor, which are eigenfunctions of L^2 and S^2 with the same eigenvalues throughout, and which are eigenfunctions of L_z and S_z with all the possible combinations of eigenvalues of these consistent with the L^2 and S^2 eigenvalues. Thus writing in the eigenvalues of L_z and S_z , the symbol sA represents the set $sA(0, \frac{1}{2})$ and $sA(0, -\frac{1}{2})$, while pA represents $pA(1, \frac{1}{2}), pA(0, \frac{1}{2}), pA(-1, \frac{1}{2}), pA(1, -\frac{1}{2})$, etc. There is a possibility of arbitrary choice of phase constants not determined by these definitions. However, we shall follow the whole treatment of Condon & Shortley in this section, and the phases will be taken as explicitly postulated there.

We shall follow the usual convention of notation and represent a product of functions in which the arguments of the first function are x_1, y_1, z_1, v_1 , those of the second x_2 , etc., by the product of the functions with all arguments omitted.

The co-detors used in the present calculation are the antisymmetric functions formed from the following types of vector coupled functions:

$$\begin{aligned} & (sAsA)^1S (sBsB)^1S, \quad (sAsA)^1S (sBsC)^1S, \\ & (sAsA)^1S (pApA)^1S, \quad (sAsB)^1S (pApA)^1S, \end{aligned} \quad (3)$$

where 1S is used in accordance with ordinary theory of spectra notation to denote linear combinations with which are eigenfunctions of L^2 and S^2 with zero eigenvalues. To derive antisymmetric functions we operate on these with

$$\mathcal{S} \equiv \sum_u \sigma_u P_u / \sqrt{(N!)},$$

where this represents a sum of all permutation operators acting on the suffixes of the variables x_1, x_2, x_3, x_i , etc., σ_u denotes -1 or $+1$ according as the permutation is odd or even, and N is the number of electrons. If we use $\mathcal{A}$ to represent this operation followed by normalization of the whole function we obtain

$$\left. \begin{aligned} \mathcal{A}(sAsA)^1S(sBsB)^1S &= \mathcal{S}sA(0, \tfrac{1}{2})sA(0, -\tfrac{1}{2})sB(0, \tfrac{1}{2})sB(0, -\tfrac{1}{2}), \\ \mathcal{A}(sAsA)^1S(sBsC)^1S &= (1/\sqrt{2})\mathcal{S}\{sA(0, \tfrac{1}{2})sA(0, -\tfrac{1}{2}) \\ &\quad \times \{sB(0, \tfrac{1}{2})sC(0, -\tfrac{1}{2}) - sB(0, -\tfrac{1}{2})sC(0, \tfrac{1}{2})\}, \\ \mathcal{A}(sAsA)^1S(pApA)^1S &= (1/\sqrt{3})\mathcal{S}\{sA(0, \tfrac{1}{2})sA(0, -\tfrac{1}{2})\}\{pA(1, \tfrac{1}{2})pA(-1, -\tfrac{1}{2}) \\ &\quad - pA(0, \tfrac{1}{2})pA(0, -\tfrac{1}{2}) + pA(-1, \tfrac{1}{2})pA(1, -\tfrac{1}{2})\}, \\ \mathcal{A}(sAsB)^1S(pApA)^1S &= (1/\sqrt{6})\mathcal{S}\{sA(0, \tfrac{1}{2})sB(0, -\tfrac{1}{2}) - sA(0, -\tfrac{1}{2})sB(0, \tfrac{1}{2})\} \\ &\quad \times \{pA(1, \tfrac{1}{2})pA(-1, -\tfrac{1}{2}) - pA(0, \tfrac{1}{2})pA(0, -\tfrac{1}{2}) + pA(-1, \tfrac{1}{2})pA(1, -\tfrac{1}{2})\}. \end{aligned} \right\} \quad (4)$$

These expressions have been obtained by the insertion of the numerical coefficients of the 1S couplings. Since $\mathcal{S}$ operating on a single normalized product gives just a single normalized detor we have explicit co-detor expansions. These are eigenfunctions of L^2 and S^2 with zero eigenvalues, and these together with the others formed interchanging sA, sB and sC in all possible ways are just the co-detors used in this calculation.

The evaluation of the integrals of the Schrödinger-Hamiltonian between detors formed from single-electron functions which are eigenfunctions of L^2 and S^2 is well known (see Condon & Shortley 1935, p. 170), and the integrals for the co-detors can easily be obtained by evaluation of the individual detor integrals. Short methods are easily obtained, but as quite an extensive theory of new reduction methods will be given in a later contribution no further details need be given here, as the direct expansions are not prohibitively long in this case. We give below explicit formulae for examples of all the different types of matrix element which occur for these co-detors. We first define the following notations for integrals over the radial factors of the functions sA, sB , etc. We use ϕ_r to represent the particular sets, thus $\phi_1 = sA, \phi_2 = sB$, etc., and $\bar{\phi}_r$ to represent the radial factor multiplying the spherical harmonic and spin function. When the eigenvalues of ϕ_r and ϕ_s for L^2 are both $L(L+1)$ we use the following notations, the corresponding quantities not occurring when the eigenvalues of L^2 are different:

$$\left. \begin{aligned} [\phi_s | \phi_t] &\equiv \int_0^\infty \bar{\phi}_s^* \bar{\phi}_t r^2 dr, \\ \left[\phi_s \left| \frac{1}{r} \right| \phi_t \right] &\equiv \int_0^\infty \bar{\phi}_s^* \bar{\phi}_t r dr, \\ [\phi_s | K | \phi_t] &\equiv \int_0^\infty \bar{\phi}_s^* \left\{ \left(-\frac{1}{2} \frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{1(1+1)}{2r} \right) \right\} \bar{\phi}_t r^2 dr. \end{aligned} \right\} \quad (5)$$

We also require the integral

$$\begin{aligned} [\phi_s \phi_t | \phi_u \phi_v]^k &\equiv \left\{ \int_0^\infty r_1^{1-k} dr_1 \int_0^{r_1} r_2^{k+2} dr_2 + \int_0^\infty r_2^{1-k} dr_2 \int_0^{r_2} r_1^{k+2} dr_1 \right\} \\ &\quad \times \{ \bar{\phi}_s^*(r_1) \bar{\phi}_t(r_1) \bar{\phi}_u^*(r_2) \bar{\phi}_v(r_2) \}, \end{aligned} \quad (6)$$

which can occur when the eigenvalues of L^2 are different. k is an integer. Then working with the methods described in Condon & Shortley (1935, p. 174) we can obtain the following values for the integrals of the Schrödinger-Hamiltonian between the different types of co-detors:

$$\begin{aligned}
 & (\mathcal{A}(sAsA)^1S (sBsB)^1S | H | \mathcal{A}(sAsA)^1S (sBsB)^1S) \\
 & = [sAsA | sAsA]^0 + [sBsB | sBsB]^0 + 4[sAsA | sBsB]^0 - 2[sAsB | sAsB]^0 \\
 & \quad - 2[sA | ZV - K | sA] - 2[sB | ZV - K | sB], \\
 & \mathcal{A}(sAsA)^1S (sBsC)^1S | H | \mathcal{A}(sAsA)^1S (sBsC)^1S) \\
 & = [sAsA | sAsA]^0 + [sBsB | sCsC]^0 + 2[sAsA | sBsB]^0 \\
 & \quad + 2[sAsA | sCsC]^0 + [sBsC | sCsB]^0 - [sAsC | sCsA]^0 - [sAsB | sBsA]^0 \\
 & \quad - 2[sA | ZV - K | sA] - [sB | ZV - K | sB] - [sC | ZV - K | sC], \\
 & (\mathcal{A}(sAsA)^1S (pApA)^1S | H | \mathcal{A}(sAsA)^1S (pApA)^1S) \\
 & = [sAsA | sAsA]^0 + [pApA | pApA]^0 + (\frac{2}{5})[pApA | pApA]^2 + 4[sAsA | pApA]^0 \\
 & \quad - \frac{2}{3}[sApA | pAsA]^1 - 2[sA | ZV - K | sA] - 2[pA | ZV - K | pA], \\
 & (\mathcal{A}(sAsB)^1S (pApA)^1S | H | \mathcal{A}(sAsB)^1S (pApA)^1S) \\
 & = [sAsA | sBsB]^0 + [pApA | pApA]^0 + \frac{2}{5}[pApA | pApA]^2 \\
 & \quad + 2[sAsA | pApA]^0 + 2[sBsB | pApA]^0 - \frac{1}{3}[sApA | pAsA] - (\frac{1}{3})[sBpA | pAsB] \\
 & \quad - [sA | ZV - K | sA] - [sB | ZV - K | sB] - 2[pA | ZV - K | pA], \\
 & (\mathcal{A}(sAsA)^1S (sBsB)^1S | H | \mathcal{A}(sAsA)^1S (sBsC)^1S) \\
 & = \sqrt{2}\{[sBsB | sBsC]^0 + 2[sAsA | sBsC]^0 - [sAsC | sBsA]^0 - [sB | ZV - K | sC]\}, \\
 & (\mathcal{A}(sAsA)^1S (sBsB)^1S | H | \mathcal{A}(sAsA)^1S (sCsC)^1S) \\
 & = [sBsC | sBsC]^0, \\
 & (\mathcal{A}(sAsA)^1S (sBsB)^1S | H | \mathcal{A}(sAsB)^1S (sCsC)^1S) \\
 & = -\sqrt{2}[sAsC | sBsC], \\
 & (\mathcal{A}(sAsB)^1S (sCsC)^1S | H | \mathcal{A}(sAsC)^1S (sBsB)^1S) \\
 & = 2[sAsB | sCsA]^0 - [sCsC | sCsB]^0 - [sBsB | sCsB]^0 \\
 & \quad - [sAsA | sCsB]^0 - [sC | ZV - K | sB], \\
 & (\mathcal{A}(sAsA)^1S (sBsB)^1S | H | \mathcal{A}(sAsA)^1S (pApA)^1S) \\
 & = -(1/\sqrt{3})[sBpA | sBpA]^1, \\
 & (\mathcal{A}(sAsA)^1S (sBsB)^1S | H | \mathcal{A}(sAsB)^1S (pApA)^1S) \\
 & = \sqrt{(\frac{2}{3})}[sApA | sBpA]^1, \\
 & (\mathcal{A}(sAsA)^1S (sBsC)^1S | H | \mathcal{A}(sAsA)^1S (pApA)^1S) \\
 & = -\sqrt{\frac{2}{3}}[sBpA | sCpA]^1, \\
 & (\mathcal{A}(sAsB)^1S (sCsC)^1S | H | \mathcal{A}(sAsB)^1S (pApA)^1S) \\
 & = -\sqrt{\frac{1}{3}}[sCpA | sCpA]^1, \\
 & (\mathcal{A}(sAsC)^1S (sBsB)^1S | H | \mathcal{A}(sAsB)^1S (pApA)^1S) \\
 & = \sqrt{\frac{1}{3}}[sCpA | sBpA]^1, \\
 & \mathcal{A}(sAsA)^1S (pApA)^1S | H | \mathcal{A}(sAsB)^1S (pApA)^1S) \\
 & = \sqrt{2}\{[sAsA | sAsB]^0 + 2[sAsB | pApA]^0 \\
 & \quad - \frac{1}{3}[sApA | pAsB]^1 - [sA | ZV - K | sB]\},
 \end{aligned}
 \tag{7}$$

These formulae show examples of all the different types of integrals which occur in the present problem with the exception of two cases where the two co-detors differ in more than two of the ϕ_s from which they are constructed, when the integrals are zero. All other cases which are not given explicitly here can be derived from the examples given by suitable interchanges of sA , sB and sC .

3. THE EVALUATION OF THE ELEMENTARY INTEGRALS

We require the values of the integrals between the sA , sB , sC , pA functions. These are linear combinations of functions which all have radial factors of the general form

$$r^{A+l} e^{-ar},$$

where l is the degree of the spherical harmonic which the factor multiplies, where A is an integer and a is any positive number. The values of the integrals for such types follow directly. Writing η_A , η_B , etc., for the sets of functions with constants (A, a) , (B, b) , etc., we have

$$\begin{aligned} [\eta_A | \eta_B] &= \int_0^\infty r^{2+A+B+2l} e^{-(a+b)r} dr \\ &= T(A+B+2l+2, a+b), \end{aligned} \quad (8)$$

where

$$T(x, y) \equiv x! / y^{x+1}. \quad (9)$$

Similarly, we have

$$[\eta_A | V | \eta_B] = \int_0^\infty r^{A+B+2l+1} e^{-(a+b)r} dr = T(A+B+2l+1, a+b), \quad (10)$$

and using a certain amount of manipulation to obtain a symmetrical form we have

$$\begin{aligned} [\eta_A | K | \eta_B] &= \frac{1}{2} \int_0^\infty r^2 dr \{ r^{A+l} e^{-ar} \} \left\{ \left(-\frac{1}{r^2} \frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} + \frac{l(l+1)}{r^2} \right) r^{B+l} e^{-br} \right\} \\ &= \frac{1}{2} \{ 1 - b^2 T(A+B+2l+2, a+b) + 2b(B+l+1) T(A+B+2l+1, a+b) \\ &\quad + [l(l+1) - (B+1)(B+l+1)] T(A+B+2l, a+b) \} \\ &= \frac{1}{2} \{ ab T(A+B+2l+2, a+b) - (Ab + Ba) T(A+B+2l+1, a+b) \\ &\quad + AB T(A+B+2l, a+b) \}. \end{aligned} \quad (11)$$

For the last type of integral in which the values of l are not necessarily the same and are denoted by l_A , l_B , etc., we obtain

$$\begin{aligned} [\eta_A \eta_B | \eta_C \eta_D]^k &= \int_0^\infty dr_1 \int_0^{r_1} dr_2 \{ r_1^{x+1-k} r_2^{y+2+k} e^{-ur_1-vr_2} \} \\ &\quad + \int_0^\infty dr_2 \int_0^{r_2} dr_1 \{ r_2^{y+1-k} r_1^{x+2+k} e^{-ur_1-vr_2} \}, \end{aligned} \quad (12)$$

where we have used

$$\left. \begin{aligned} a+b &= u, & c+d &= v, \\ A+B+l_A+l_B &= x, & C+D+l_C+l_D &= y. \end{aligned} \right\} \quad (13)$$

It is convenient to define

$$U(p, q, u, v) \equiv \int_0^\infty dr_1 \int_0^{r_1} dr_2 \{r_1^p r_2^q e^{-ur_1 - vr_2}\} \\ = \left(-\frac{\partial}{\partial u}\right)^p \left(-\frac{\partial}{\partial v}\right)^q \frac{1}{u(u+v)} = \sum_{r=0}^p \frac{p!(p+q-r)!}{(p-r)!} \frac{1}{u^{1+r}} \frac{1}{(u+v)^{p+q+1-r}}. \quad (14)$$

Hence we can write

$$[\eta_A \eta_B | \eta_C \eta_D]^k = U(x+1-k, y+2+k, u, v) + U(y+1-k, x+2+k, v, u). \quad (15)$$

We have thus evaluated all the necessary integrals and have obtained forms which are fairly simple for the systematic numerical evaluation of considerable numbers of these.

4 THE EVALUATION OF THE INTEGRALS BETWEEN THE ORTHOGONAL SINGLE-ELECTRON FUNCTIONS

The procedure by which the sets of orthogonal functions sA , sB , sC and pA were obtained was to take four radial functions, multiply these by the spherical harmonic, make orthogonal by the Schmidt process, and multiply by the α and β spin functions. The particular radial functions used in this calculation were e^{-4r} , re^{-r} , e^{-3r} for the s functions and re^{-r} for the pA functions.

The reason why these particular functions were chosen was that e^{-4r} and re^{-r} are approximately the functions used by Duncanson & Coulson (1944) for the best approximation to the beryllium wave function when only a single co-deter term based on radial exponential terms was used. The reason for including the e^{-3r} function was that a function of this type had been found very effective in lowering the calculated energy in the case of a preliminary trial for a He-like ion. Detailed inspection of the way the pA function lowers the calculated energy suggests that this will be large when the pA function overlaps the sB function strongly, and this is approximately a maximum if we choose re^{-r} . Theoretically we ought to compare the quality of different choices of initial functions but since these empirical reasons have enabled such a satisfactory result to be obtained this was not necessary.

The reason why the whole calculation is carried out in terms of orthogonal linear combinations of these initial functions instead of the initial functions themselves is purely a matter of mathematical convenience, and has nothing to do with the physical theory. Actually, it can easily be proved that if we use all the possible Slater determinants from a given set of initial functions the final answer is independent of what particular linear combinations are used, but if non-orthogonal combinations were used the formulae for the integrals between the co-deters would become impossibly complicated. Hence we use a set of orthogonal linear combinations and choose it arbitrarily for the greatest detailed convenience.

When we combine these radial functions with the spherical harmonics we obtain the functions which we denote in general by η_i and in particular by

$$\left. \begin{aligned} sa &= e^{-4r} S^{00}, & sb &= re^{-r} S^{00}, & sc &= e^{-3r} S^{00} \\ pa(1) &= re^{-r} S^{11}, & pa(0) &= re^{-r} S^{10}, & pa(1) &= re^{-r} S^{1,-1}, \end{aligned} \right\} \quad (16)$$

where S^{lm} represents the (l, m) spherical harmonic with the same conventions as used by Condon & Shortley. The pa functions are automatically orthogonal to each other and the s function, and need only be normalized. To obtain orthogonal combinations of the sa, sb, sc we used the ordinary Schmidt process giving functions ϕ_r which satisfy

$$[\phi_r | \phi_s] = \sum X_i^r X_j^s [\eta_i | \eta_j] = \delta_{rs}. \quad (17)$$

The radial factors obtained were

$$\left. \begin{aligned} \overline{sA} &= 16\overline{sa}, \quad \overline{sB} = -2.883, 508, 0\overline{sa} + 1.173, 302, 4\overline{sb}, \\ \overline{sC} &= -69.597, 017\overline{sa} - 0.596, 885, 7\overline{sb} + 47.607, 282\overline{sc}, \\ \overline{pA} &= 1.154, 700, 5\overline{pa}. \end{aligned} \right\} \quad (18)$$

The complete single-electron functions used to construct the detors were thus

$$\left. \begin{aligned} sA(0, \tfrac{1}{2}) &= \overline{sA} S^{00} \mu(\tfrac{1}{2}), \quad \overline{sA}(0, -\tfrac{1}{2}) = sA S^{00} \mu(-\tfrac{1}{2}), \\ pA(1, \tfrac{1}{2}) &= \overline{pA} S^{11} \mu(\tfrac{1}{2}), \text{ etc.} \end{aligned} \right\} \quad (19)$$

where $\mu(\tfrac{1}{2})$ and $\mu(-\tfrac{1}{2})$ are the spin functions which are eigenfunctions of s_z , or α and β as they are frequently represented.

We have shown how to evaluate the integrals between the above non-orthogonal functions, and we now require the integrals between these orthogonal linear combinations. The cases $[sA | K | sB]$ and $[sAsB | pApA]^0$ will serve as characteristic examples. These are given immediately by the relations

$$\left. \begin{aligned} [sA | K | sB] &= \sum_{r,s} X_r^{sA} X_s^{sB} [\eta_r | K | \eta_s], \\ [sAsB | pApA]^0 &= \sum_{r,s,t,u} X_r^{sA} X_s^{sB} X_t^{pA} X_u^{pA} [\eta_r \eta_s | \eta_t \eta_u]^0, \end{aligned} \right\} \quad (20)$$

where X_r^{sA} , etc., denote the numerical coefficients of equations (18) and η_r the sa, sb, sc, pa functions.

The evaluation of the first integral is simple, the matrix $[\eta_r | K | \eta_s]$ is written in square form and contracted with the column X_r^{sA} to give the matrix $[sA | K | \eta_s]$ which is then contracted with regard to the untransformed suffix with the column X_s^{sB} to give $[sA | K | sB]$.

For the second integral the situation is more complicated, since we are dealing with the numerical transformation of a four-suffix matrix. The method employed was to write $[\eta_r \eta_s | \eta_t \eta_u]^0$ as a two-suffix matrix $[p | M | q]^0$, that is, with each row and column corresponding to a pair of values r and s . The corresponding vectors

$$\left. \begin{aligned} T_p^{sAsB} &= X_r^{sA} X_s^{sB} + X_s^{sA} X_r^{sB}, \quad \text{when } r \neq s \\ &= X_r^{sA} X_r^{sB}, \quad \text{when } r = s, \end{aligned} \right\} \quad (21)$$

were then formed, and by two contractions of these with $[p | M | q]^0$ each integral of the type $[sAsB | pApA]^0$ was found. It is this transformation which is one of the heaviest portions of the numerical calculation.

It is then only necessary to insert these values in the formulae of § 2 to give the numerical values for the final H_{rs} matrix, the solution of which will be discussed in the next section.

5 THE NUMERICAL SOLUTION OF THE EIGENVECTOR PROBLEM

We require to find by the Ritz variational method a linear combination

$$\Psi = \sum_r Y_r \Phi_r, \quad (22)$$

where the Φ_r are the co-detors of § 2, such that Ψ will be the best approximation to the solution of the exact problem

$$(H - E)\psi = 0. \quad (23)$$

For this it is well known that the coefficients Y_r and E must satisfy

$$\sum_s H_{rs} Y_s - E Y_r = 0, \quad (24)$$

where E is the energy corresponding to this approximation.

We know the numerical values of H_{rs} , and we have a numerical eigenvector problem to solve. General methods of solving such a problem are known but a much easier method is available in the present circumstances. One of the terms is a reasonable approximation to the true Ψ , so that the coefficient of this term in Ψ is much larger than the other coefficients. We shall for convenience put this equal to unity since if we find an unnormalized solution we can normalize it at the last stage after the solution of the eigenvector problem has been carried through to the desired accuracy. We shall put this term to be Φ_1 and use $\bar{E}$ to denote $E - H_{11}$. Hence the Y_r satisfy the equivalent equations

$$\sum_s \bar{H}_{rs} Y_s - \bar{E} Y_r = 0, \quad (25)$$

where $\bar{H}_{rr} = H_{rr} - H_{11}$ and $\bar{H}_{rs} = H_{rs}$ when $r \neq s$. Now in this particular problem the diagonal coefficients are large compared with the other coefficients, and we can use a very simple iterative process for solution. Consider an approximate result Y_r^0 and let δY_r represent the corrections to these to give the true answer. Let us take an approximation to $\bar{E}$ to satisfy the first equation, that is,

$$\bar{E}_0 = \sum_s \bar{H}_{1s} Y_s^0. \quad (26)$$

Then in the rest of the equations we neglect the terms which are quadratic in δY , δE and H_{rs} as negligible compared with the linear terms $\bar{H}_{rr} \delta Y_r$, so that

$$\sum_s \bar{H}_{rs} Y_s^0 - \bar{E}_0 Y_r + \sum_r \bar{H}_{rr} \delta Y_r = 0 \quad \text{and} \quad \delta \bar{E} = 0, \quad (27)$$

which gives us the values of the corrections as

$$\left. \begin{aligned} \delta Y_r &= (-1 / \bar{H}_{rr}) \Delta_r, \\ \Delta_r &\equiv \sum_s \bar{H}_{rs} Y_s^0 - \bar{E}_0 Y_r. \end{aligned} \right\} \quad (28)$$

where

The quantities Δ_r should in an exact solution be zero, and this method of improvement of successive solutions is continued until the Δ_r are judged to be small enough. The convergence was extremely rapid, so that effectively commencing with (1, 0, 0, 0, ...) as the first approximation gave a solution which was sufficiently

accurate after three or four iterations. It may be noted, however, that the value of $\bar{E}$ quoted for any approximation should be calculated from

$$\bar{E} = \sum_{r,s} Y_r \bar{H}_{rs} Y_s / \sum_r Y_r^2. \quad (29)$$

This expression is in strict accordance with the variational principle and is much less sensitive to the residual errors in the Y_r than the value E_0 obtained in the iterative process.

It is very probable that this approximation process will be applicable to most problems of this type if a certain amount of care is taken in choosing the initial functions. If the Φ_r used correspond to approximate solutions of the eigenvector problem the method will certainly work, since the interaction terms will be small compared with the diagonal terms. This is equivalent to saying the Φ_r should be approximately excited states of the atom within the limited region of configuration space to which our approximation of taking only a finite number of elementary terms corresponds. The fact that one function is an approximation to the ground state is a step towards this, but apparently we cannot make any general statement on the possibility of arranging Φ_r to be of this type in other cases.

The solution of the eigenvector problem without this method would not be prohibitively laborious, but we should lose considerable useful flexibility, since in the present case we are able to estimate by inspection the effect of including or omitting particular co-determinants.

6. NUMERICAL VALUES

For accidental reasons practically the whole calculation was carried out twice using rather different formulations, but at each step at least one of the calculations corresponded very closely to the explicit formulae of the previous sections. The formulae of § 2 were first worked out on a non-normalized basis differing from those stated by factors of $\sqrt{2}$, $\sqrt{3}$ or $\sqrt{6}$. The calculation was first carried out in accordance with this. Later it was found that this basis was not as applicable to more complicated atoms, and the formulae as given were calculated by another theoretical method. Other differences were made in the numerical procedure, and it was judged best to transform the first figures for H_{rs} to the new basis and compare them with the recalculated values. It appears that this effectively eliminates the possibility of any error. Actually the orthogonalization process was not repeated, but the calculation of the $[sA | sB]$ quantities was carried out on two different bases and these found to be 0 or 1 as required.

All initial quantities were evaluated to seven significant figures or over, and the various manipulations all performed to this accuracy. This causes the accuracy of the most inaccurate of the $[\phi_r \phi_s | \phi_t \phi_u]$ quantities to be about four significant figures, since the process of taking orthogonal combinations involves considerable differencing. This accuracy is, of course, well beyond the requirements consistent with the errors involved by taking only a finite number of terms to approximate to the wave function and very much beyond the accuracy of evaluation of integrals in the Hartree method. Actually the accuracy is much beyond immediate require-

ments, as the more inaccurate quantities only occur with very small coefficients, but this accuracy is ready for further improvements by the inclusion of more terms, and it actually facilitates checking in such processes as matrix contraction.

The eigenvector problem was solved to two degrees of approximation, once using the six functions which inspection showed to be most effective in giving a low energy, and once using ten functions. This last case really corresponds to using all the functions which can be made from the set of initial single-electron functions, since those omitted were judged to be extremely unlikely to give improvements for detailed empirical reasons. It will be seen that actually the ten functions gave no significant improvement on the energy of the six functions. The accuracy to which the iteration process was carried was to make all the Δ_r less than 0.005. The solutions and the energies are given in table 1, and the values of the $\bar{H}_{rs}$ for the six-function case in table 2. The Y_r are given to four decimal places, as adjustment to the last place was necessary in some cases to satisfy the Δ_r condition, but this is well beyond the physical accuracy of the solution. It is difficult to make any precise judgement on what is the real physical accuracy of the approximation.

TABLE 1. WAVE FUNCTIONS AND ENERGIES

	Y_r	Y'_r
$\mathcal{A}(sAsA)^1S (sBsB)^1S$	1.0	1.0
$\mathcal{A}(sBsB)^1S (sCsC)^1S$	-0.0200	-0.0200
$\mathcal{A}(sAsA)^1S (sBsC)^1S$	-0.0509	-0.0500
$\mathcal{A}(sAsC)^1S (sBsB)^1S$	0.1055	0.1060
$\mathcal{A}(sAsA)^1S (pApA)^1S$	0.3188	0.3246
$\mathcal{A}(sAsC)^1S (pApA)^1S$	0.0343	0.0348
$\mathcal{A}(sAsA)^1S (sCsC)^1S$	—	-0.0022
$\mathcal{A}(sAsB)^1S (sCsC)^1S$	—	-0.0040
$\mathcal{A}(sBsB)^1S (pApA)^1S$	—	0.0009
$\mathcal{A}(sAsB)^1S (pApA)^1S$	—	0.0124
Calculated E	-14.6211	-14.6220
Experimental E	-14.66	
Hartree E	-14.58	

TABLE 2. $\bar{H}_{rs}$

	$sAsA \ sBsB$	$sBsB \ sCsC$	$sAsA \ sBsC$	$sAsC \ sBsB$	$sAsA \ pApA$	$sAsC \ pApA$
$sAsA \ sBsB$	0.0	0.3967	0.2488	-0.9416	-0.1361	0.0
$sBsB \ sCsC$	—	17.0628	-0.1257	-0.5601	0.0	0.0
$sAsA \ sBsC$	—	—	5.0387	0.1499	-0.0099	0.0196
$sAsC \ sBsB$	$(\bar{H}_{rs} = \bar{H}_{sr}$ and $H_{11} = -14.4577)$			8.7275	0.0	-0.1361
$sAsA \ pApA$				—	0.3677	-0.9502
$sAsC \ pApA$	—	—	—	—	—	9.1242

The Hartree value of the total energy given in the table was obtained by taking the value for Be^{++} from Hartree & Hartree (1935) and adding on the value for the $2s^2$ shell from Hartree & Hartree (1936). These authors give the value -14.56 in the first paper, but the combination as taken is a lower value, and the above comparison shows that our value surpasses this most favourable combination to the extent of halving the deviation from the experimental value.

7. DISCUSSION

The results of the calculation are gratifying, since apparently less numerical work was performed than in the calculation of a Hartree-Fock wave function, but according to the minimization of energy test the result has moved half-way towards the experimental result from the Hartree result.

The wave function found is in terms of simple analytical functions, and although the mathematical theory for the calculation of atomic properties from it is more complicated than the calculation from a Hartree type the numerical work is much less since numerical integrations are not necessary. It will be shown in subsequent communications that the mathematical theory actually can be made extremely simple when a few advanced theorems are established.

The general method of the calculation is a true converging process, so that any degree of accuracy can be achieved by the inclusion of sufficient terms without in any way altering the general character of the calculation. This is not true for the Hartree-Fock method, although it is true for the superposition of configurations method as applied by Hartree *et al.* (1939). However, these workers only found it practicable to include two independent co-deter terms, and apparently the numerical work was considerable. In our case the theoretical part of the calculation is more complicated, the difficult part being the theoretical calculation of the formulae for the H_{rs} elements, but the numerical work is apparently considerably less. This mathematical work becomes more involved for heavier atoms and excited states, but a series of theorems which provide a systematic method for treating these has been worked out and will be communicated later.

One misconception might arise with regard to the significance of the method. It might be thought that the different co-detors used corresponded approximately to other discrete excited states of the atom. This is not true even approximately. If we tried to trace such a correspondence we should find that some of the co-detors, very effective in reducing the energy, correspond even more to the continuous range of states, that is to a partially ionized atom, than to the lower discrete states. The correct way of regarding the method is that we are using a system of functions, complete in the mathematical sense so that we can converge to the accurate wave function. It is known that the wave functions of the discrete levels do not constitute such a complete system, and hence our system must include the continuous range of states as well.

Although beryllium only contains four electrons there is no aspect of the calculation which depends on this, and there is every reason to hope that this method will be practically applicable to most atoms and their excited states. It might, however, be recalled that this investigation was made partly as a contribution towards the development of practical methods for molecules as well as atoms. The method we have suggested for molecules in the preceding paper has much in common with this method, except that other functions are used and the evaluation of the corresponding integrals is much more laborious. This prior treatment of atoms has helped to improve the detailed numerical techniques for the common parts of these calculations and should reduce the practical exploratory work for the more complicated molecular case.

The author is indebted to the University of London for the award of an I.C.I. Research Fellowship, during the tenure of which this investigation was carried out.

REFERENCES

- Boys, S. F. 1950 *Proc. Roy. Soc. A*, **200**, 542.
 Condon, E. U. & Shortley, G. H. 1935 *Theory of atomic spectra*. Cambridge University Press.
 Duncanson, W. E. & Coulson, C. A. 1944 *Proc. Roy. Soc. Edinb.* **62**, 37.
 Hartree, D. R. & Hartree, W. 1935 *Proc. Roy. Soc. A*, **150**, 9.
 Hartree, D. R. & Hartree, W. 1936 *Proc. Roy. Soc. A*, **154**, 588.
 Hartree, D. R., Hartree, W. & Swirles, B. 1939 *Phil. Trans. A*, **238**, 229.

The dispersion of a pressure pulse in the atmosphere

BY R. S. SCORER, *Corpus Christi College, University of Cambridge*

(Communicated by Sir Geoffrey Taylor, F.R.S.—Received 31 August 1949—
 Revised 7 November 1949)

The object is to find what pressure oscillations would be observed on the ground at a great distance from an explosion. The explosion is represented mathematically by a Fourier integral, corresponding to the introduction of a large volume into the atmosphere at a point on the ground. The resulting pulse is calculated for various distances for a model atmosphere consisting of a troposphere with a constant lapse-rate of temperature and an isothermal stratosphere. It is composed of those oscillations that can be propagated horizontally as gravity waves in this model atmosphere, namely, those of period exceeding a cut-off period of 111 sec. The pulse consists of a series of waves of decreasing amplitude and period, terminating with a period of 12.7 sec.

The results are compared with the oscillations observed on the occasion of the fall of the Great Siberian Meteorite and the energy which it is estimated to have communicated to the atmosphere is about 4×10^{24} ergs only a fraction of which resided in the gravity wave. Neglect of the warmer layers in the higher levels in the stratosphere means that the calculated pulse terminated too soon, and a second series of waves of considerable amplitude and of greater frequency is completely absent. The form of these has not been calculated because of the prohibitive amount of computing involved.

1. INTRODUCTION

The problem we seek to solve is to find what pressure oscillations would be observed at a great distance from a large explosion at the earth's surface. The amplitude and form of the oscillations, when related to the magnitude and distance of the explosion, would then provide, on comparison with observations, an estimate of the energy communicated to the atmosphere by the Great Siberian Meteorite of 1908 or the eruption of Krakatoa in 1883.

In order to render the problem reasonably simple many assumptions are made. The earth is assumed to be spherical and the atmosphere to be almost horizontally stratified—the temperature and lapse-rate of temperature and the height of the tropopause are therefore assumed to be uniform over the earth's surface and the

stratosphere to be isothermal. There is assumed to be no motion other than that produced by the explosion and this is supposed to take place adiabatically and to be of small magnitude. The actual atmosphere is thus not very faithfully simulated, but apart from the wind structure, which is too complex to represent anyway, it is apparent from the equations that no great difference would result from these assumptions except from the neglect of the warmer layers of the stratosphere above about 30 km. These are known, from the work of Whipple (1935) and others, to reflect audible air waves downwards, and this fact is discussed when theory and observation are compared. These audible air waves were treated in terms of geometrical optics, the velocity of propagation being dependent only on the temperature at any point; the oscillations to be discussed here, however, are primarily gravity waves of much longer period than sound waves.

The explosion causes a localized upward displacement of the surfaces of constant density which were originally horizontal, and sends out circular wave fronts in the same way as the intrusion of an object into the surface of static water gives rise to a local elevation of the surface followed by a set of circular waves travelling outwards. The atmosphere is a dispersive medium: the phase velocity depends upon the wavelength; and so it is necessary to discover first the relationship between them. That such a relationship exists depends upon the fact that the horizontal and vertical variations of pressure (or whatever quantity is chosen to represent the disturbance) may be separated mathematically, and vertical wave fronts may be propagated horizontally over the earth's surface, the amplitude and phase of the wave at any height being determined by a differential equation relating the pressure to the height only. To this equation the boundary conditions at the 'top' and bottom of the atmosphere are applied, and when this is done the required relationship between phase velocity and frequency is obtained.

It then only remains to synthesize the explosion by means of a Fourier integral. This is in effect a boundary condition applied to the two differential equations for the vertical and horizontal variations of pressure, and it asserts that over a small region the ground rises rapidly for a short time and introduces a calculated volume into the atmosphere. The surfaces of constant density are suddenly distorted upwards and a 'pulse' moves away radially. The waves of various frequencies composing the pulse are dispersed and the pulse form changes as it travels out; the pulse is observed by the pressure variations it produces at the ground, and it is these that have been calculated for various distances from the explosion.

Taylor (1936) has considered the propagation of long-period waves in connexion with the Krakatoa air wave; Pekeris (1937) and Wilkes & Weekes (1947) discussed the semidiurnal pressure oscillations in terms of long waves; and later Pekeris (1948) discussed the propagation of a pulse taking all frequencies into account, but the analysis and numerical work were not pursued to predict the exact form of the observed pulse. The present method of approach and derivation of the equations, though it arose out of a previous investigation (Scorer 1949), is fundamentally the same as that of Pekeris in the use of Fourier integrals, but differs in detail and technique. Finally, a new function had to be tabulated (Scorer 1950) before the integral giving the pulse could be evaluated.

2. DERIVATION OF THE DIFFERENTIAL EQUATIONS

The values of quantities in the undisturbed atmosphere are denoted by suffix 0, the values at the ground by suffix 1, and at the tropopause by suffix 2. Plain letters are used for the disturbance; thus p_{01} is the undisturbed pressure at the ground, ρ is the disturbance of density, etc., $\rho_0 + \rho$ = density, $p_0 + p$ = pressure, $T_0 + T$ = absolute temperature, $\tau_0 + \tau = \frac{1}{T_0 + T} \left(\frac{p_0 + p}{p_{01}} \right)^{(\gamma-1)/\gamma}$ = reciprocal of potential temperature, $\varpi_0 + \varpi = \frac{\gamma}{\gamma-1} R \left(\frac{p_0 + p}{p_{01}} \right)^{(\gamma-1)/\gamma}$ = 'modified pressure', where R is the gas constant, and γ is the ratio of the specific heats of air. θ, ϕ, z are the spherical polar co-ordinates (see figure 1) of the point where the velocity is u, v, w . The atmosphere is assumed to be shallow so that a , the earth's radius, is written for $z + a$, z being the vertical co-ordinate measured from a level not far from the ground. In this co-ordinate system

$$\text{grad} = \frac{\partial}{a \partial \theta}, \quad \frac{\partial}{a \sin \theta \partial \phi}, \quad \frac{\partial}{\partial z},$$

$$\text{div} (u, v, w) = \frac{\partial(u \sin \theta)}{a \sin \theta \partial \theta} + \frac{\partial v}{a \sin \theta \partial \phi} + \frac{\partial w}{\partial z}.$$

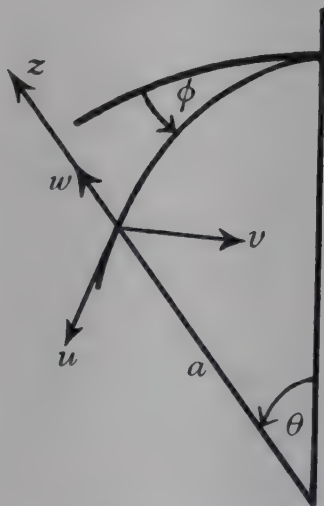


FIGURE 1. System of co-ordinates.

First, a periodic motion, with time factor $e^{i\sigma t}$, is considered; so that $\partial/\partial t = i\sigma$. Having established equations to determine the spatial variation of the 'modified pressure' for a given frequency σ , Fourier's integral theorem may be used to prescribe the time variation of the spatial boundary conditions which in turn are represented by a Fourier integral in k in a manner fully discussed later. k^2 is the constant of separation of the horizontal and vertical variations and is introduced to split up equation (10) below. The motion is assumed small so that products of the small quantities $\rho, p, T, \tau, \varpi, u, v, w$ and their derivatives are neglected. The equation of state is assumed, viz.

$$p_0 + p = (\rho_0 + \rho) R(T_0 + T).$$

If for the moment we write differentials for the disturbance we have from the definition of ϖ given above

$$\varpi = d\varpi_0 = \frac{R}{p_{01}} \left(\frac{p_0}{p_{01}} \right)^{-1/\gamma} dp_0.$$

But

$$\tau_0 = \frac{1}{T_0} \cdot \frac{p_0}{p_{01}} \left(\frac{p_0}{p_{01}} \right)^{-1/\gamma},$$

so that

$$p = dp_0 = \frac{p_0}{\tau_0 T_0 R} d\varpi_0 = \frac{\rho_0}{\tau_0} \varpi. \quad (1)$$

Since the potential temperature is invariant in adiabatic motion

$$\frac{D}{Dt} (\tau_0 + \tau) = 0,$$

where

$$\frac{D}{Dt} = (u, v, w) \text{grad} + \frac{\partial}{\partial t},$$

and for a small disturbance, assuming that the horizontal variation of τ_0 can be ignored,* this gives

$$i\sigma\tau + \tau'_0 w = 0, \quad (2)$$

where a prime is used to denote $\partial/\partial z$ when applied to τ and ϖ , and later to χ .

The effect of the earth's rotation can be shown to be very small, the burden of the proof being that a pulse takes only a fraction of a day to pass any point on the ground. If it is neglected the equation of motion is

$$\frac{D}{Dt} (u, v, w) = -\frac{1}{\rho_0 + \rho} \text{grad} (p_0 + p) + (0, 0, -g).$$

Taking the scalar product of this with (u, v, w) , for small motion we obtain, using (1),

$$\begin{aligned} \frac{1}{\rho_0 + \rho} \frac{D}{Dt} (p_0 + p) &= -gw + \frac{1}{\rho_0 + \rho} \frac{\partial}{\partial t} (p_0 + p) \\ &= -gw + \frac{i\sigma}{\tau_0} \varpi. \end{aligned} \quad (3)$$

The adiabatic equation may alternatively be written

$$\frac{D}{Dt} (p_0 + p) = c^2 \frac{D}{Dt} (\rho_0 + \rho), \quad (4)$$

where $c^2 = \gamma RT_0$; while the equation of continuity is

$$\frac{1}{\rho_0 + \rho} \frac{D}{Dt} (\rho_0 + \rho) + \text{div} (u, v, w) = 0,$$

from which $\rho_0 + \rho$ can be eliminated by (4) and then $p_0 + p$ by (3) to give

$$\frac{1}{c^2} \left(-gw + \frac{i\sigma}{\tau_0} \varpi \right) + \text{div} (u, v, w) = 0$$

or

$$\left(\frac{\cot \theta}{a} + \frac{\partial}{a \partial \theta} \right) u + \frac{\partial}{a \sin \theta \partial \phi} v + \left(\frac{\partial}{\partial z} - \frac{g}{c^2} \right) w + \frac{i\sigma}{c^2 \tau_0} \varpi = 0. \quad (5)$$

* This is justified on the grounds that we are not studying the motion due to an existing non-horizontal stratification of potential temperature but another small motion which is superposable and can be studied independently.

The quantities ϖ and τ have been so defined that

$$\frac{1}{\rho_0 + \rho} \text{grad}(p_0 + p) = \frac{1}{\tau_0 + \tau} \text{grad}(\varpi_0 + \varpi),$$

and writing this into the equation of motion, for small motion,

$$\frac{\partial}{\partial t}(u, v, w) = -\frac{1}{\tau_0 + \tau} \text{grad}(\varpi_0 + \varpi) + (0, 0, -g),$$

so that, ignoring horizontal variations of ϖ_0 , the first and second component equations are

$$i\sigma u + \frac{1}{\tau_0 a} \frac{\partial \varpi}{\partial \theta} = 0, \quad (6)$$

$$i\sigma v + \frac{1}{\tau_0 a \sin \theta} \frac{\partial \varpi}{\partial \phi} = 0, \quad (7)$$

while the third component is

$$i\sigma w + \frac{1}{\tau_0 + \tau} \frac{\partial}{\partial z}(\varpi_0 + \varpi) + g = 0,$$

or

$$i\sigma \tau_0 w + \frac{\partial \varpi_0}{\partial z} + \frac{\partial \varpi}{\partial z} + g(\tau_0 + \tau) = 0.$$

Subtracting the hydrostatic equation (see footnote to preceding page)

$$\frac{\partial \varpi_0}{\partial z} + g\tau_0 = 0,$$

and substituting for τ in terms of w by means of (2) this becomes

$$\left(i\sigma - \frac{g\tau'_0}{i\sigma\tau_0}\right)w + \frac{1}{\tau_0} \frac{\partial \varpi}{\partial z} = 0. \quad (8)$$

Equations (6), (7) and (8) are now used to eliminate u, v, w from equation (5), and the consequent equation for the 'modified pressure' is

$$\begin{aligned} \left(\frac{\cot \theta}{a} + \frac{\partial}{a \partial \theta}\right) \left(\frac{i}{\sigma a \tau_0} \frac{\partial \varpi}{\partial \theta}\right) + \frac{\partial}{a \sin \theta \partial \phi} \left(\frac{i}{\sigma a \tau_0 \sin \theta} \frac{\partial \varpi}{\partial \phi}\right) \\ + \left(\frac{\partial}{\partial z} - \frac{g}{c^2}\right) \left(\frac{i\sigma}{\sigma^2 + g\tau'_0/\tau_0} \frac{1}{\tau_0} \frac{\partial \varpi}{\partial z}\right) + \frac{i\sigma}{c^2 \tau_0} \varpi = 0. \end{aligned} \quad (9)$$

It is now assumed that the atmosphere is sufficiently nearly horizontally stratified for the horizontal variation of τ_0 to be ignored compared with that of ϖ , i.e. that it has negligible horizontal variation in the region occupied by the pulse, then on multiplying by $\sigma\tau_0/i$ equation (9) becomes

$$\left(\frac{\cot \theta}{a} + \frac{\partial}{a^2 \partial \theta}\right) \frac{\partial \varpi}{\partial \theta} + \frac{1}{a^2 \sin^2 \theta} \frac{\partial^2 \varpi}{\partial \phi^2} = -\tau_0 \left(\frac{\partial}{\partial z} - \frac{g}{c^2}\right) \left(\frac{\sigma^2}{\sigma^2 \tau_0 + g\tau'_0} \frac{\partial \varpi}{\partial z}\right) - \frac{\sigma^2}{c^2} \varpi. \quad (10)$$

The horizontal and vertical variations of ϖ are thus separated, and in the particular case of symmetry about the co-ordinate pole (which will be taken at the explosion) $\partial/\partial \phi = 0$, $v = 0$, and it is permissible to write

$$\varpi \equiv \varpi(z) \bar{\varpi}(\theta).$$

Dividing (10) by ϖ and then putting each side equal to k^2 (since the left-hand side is independent of z and the right-hand side is independent of θ) we obtain

$$\left(\cot \frac{r}{a} + \frac{\partial}{\partial r}\right) \frac{\partial \bar{\varpi}}{\partial r} = k^2 \bar{\varpi}, \quad (11)$$

$$\varpi'' - \left(\frac{g\tau_0'' + \sigma^2 \tau_0'}{g\tau_0' + \sigma^2 \tau_0} + \frac{g}{c^2}\right) \varpi' + (g\tau_0'/\tau_0 + \sigma^2) \left(\frac{1}{c^2} - \frac{k^2}{\sigma^2}\right) \varpi = 0, \quad (12)$$

where $r, = a\theta$, is the great circle distance from the pole.

This treatment differs from earlier ones (Taylor 1936; Pekeris 1937) in taking ϖ as the dependent variable in the place of the divergence of velocity. This is done because ϖ is more closely related to the disturbance of pressure (equation (1)).

If suitable boundary conditions are imposed upon ϖ , equations (11) and (12) suffice to determine ϖ throughout the whole atmosphere, and the velocity is given by (6) and (8) which reduce to

$$u = \frac{i}{\sigma\tau_0} \frac{\partial \varpi}{\partial r}, \quad (13)$$

$$w = \frac{i\sigma}{\sigma^2\tau_0 + g\tau_0'} \varpi'. \quad (14)$$

Equation (11) is independent of the way in which τ_0 varies with height. We next derive the form taken by equation (12) in the two layers of the model atmosphere we have chosen.

(a) *The troposphere.* When there is a constant lapse-rate of temperature and if the origin of z is suitably chosen

$$T_0 = -\mu \frac{(\gamma-1)g}{\gamma R} z,$$

and

$$\tau_0 = \frac{1}{T_0} \left(\frac{p_0}{p_{01}}\right)^{(\gamma-1)/\gamma},$$

$$\tau_0' = \frac{(1-\mu)\tau_0}{\mu z},$$

$$\tau_0'' = \frac{(1-\mu)(1-2\mu)\tau_0}{\mu^2 z^2},$$

where μ is a numerical constant. When $\mu = 1$ the lapse-rate is the dry adiabatic. In the troposphere $0 < \mu < 1$, and on substituting for c^2 and τ_0 and its derivatives and writing

$$\varpi = Z\chi, \quad (15)$$

equation (12) becomes

$$\chi'' = S\chi, \quad (16)$$

provided that

$$Z = z^{-\frac{1}{2}m}(z+b)^{\frac{1}{2}}, \quad (17)$$

where

$$\left. \begin{aligned} p &= -n + \frac{1}{4}(2-m), \\ n &= \frac{1-\mu}{(\gamma-1)\mu^2}, \\ m &= \frac{(2\mu-1) + 1/(\gamma-1)}{\mu}, \\ l &= -\frac{\sigma^2}{(\gamma-1)\mu g} - \frac{(1-\mu)gk^2}{\mu\sigma^2} = -l_1 - bk^2, \\ b &= \frac{(1-\mu)g}{\mu\sigma^2}, \end{aligned} \right\} \quad (18)$$

and

$$S = \frac{p}{z^2} - \frac{m/2b - l_1}{z} + \frac{\frac{3}{4}}{(z+b)^2} + \frac{m/2b}{z+b} + \left(\frac{b}{z} + 1\right)k^2. \quad (19)$$

The quantities p , n , m , l and b are determined by the lapse-rate which determines μ , by the constants γ and g , and by σ and k which determine the frequency and horizontal wave-length of the disturbance. Equation (12) has been reduced to the form (16) for purposes of numerical integration required later, so that an integration formula involving even differences of χ and χ'' only can be used.

(b) *The stratosphere.* The stratosphere is assumed to be isothermal at temperature T_s , with $c = c_s$, $\mu = 0$:

$$\tau_0 \propto \exp\left(\frac{(1-\gamma)g}{c_s^2}z\right),$$

$$\tau'_0 = -\frac{(\gamma-1)g}{c_s^2}\tau_0,$$

$$\tau''_0 = \frac{(\gamma-1)^2g^2}{c_s^4}\tau_0,$$

and equation (12) reduces to

$$\varpi'' - \frac{(2-\gamma)g}{c_s^2}\varpi' + \left(\sigma^2 - \frac{(\gamma-1)g^2}{c_s^2}\right)\left(\frac{1}{c_s^2} - \frac{k^2}{\sigma^2}\right)\varpi = 0, \quad (20)$$

and since the coefficients are constant, the solution is

$$\varpi = \alpha e^{\kappa z} + \beta e^{\lambda z},$$

$$\kappa, \lambda = \frac{(2-\gamma)g}{2c_s^2} \pm \left\{ \left(\frac{(2-\gamma)g}{2c_s^2} \right)^2 + \left(\frac{(\gamma-1)g}{c_s^2} - \sigma^2 \right) \left(\frac{1}{c_s^2} - \frac{k^2}{\sigma^2} \right) \right\}^{\frac{1}{2}}, \quad (21)$$

α and β being constants.

3. BOUNDARY CONDITIONS; FREE WAVES; THE CUT-OFF FREQUENCY

Following Pekeris, we assume that all the oscillations considered, if they are to be set up by a source of finite energy, must possess a finite kinetic energy in a vertical column of air extending infinitely upwards. If the stratosphere continues indefinitely upwards isothermally

$$\tau_0 \propto \exp\left(\frac{(1-\gamma)g}{c_s^2}z\right),$$

$$\rho_0 \propto \exp\left(-\frac{\gamma g}{c_s^2}z\right),$$

and if ϖ , and therefore also $\tau_0 u$ and $\tau_0 w$, is proportional to $e^{\kappa z}$ or $e^{\lambda z}$ then the kinetic-energy density is proportional to

$$\rho u^2 \propto \exp\left[\pm 2z\left\{\left(\frac{(2-\gamma)g}{2c_s^2}\right)^2 + \left(\frac{(\gamma-1)g^2}{c_s^2} - \sigma^2\right)\left(\frac{1}{c_s^2} - \frac{k^2}{\sigma^2}\right)\right\}^{\frac{1}{2}}\right]. \quad (22)$$

The condition permits only the negative sign, and so in the solution of equation (20) $\alpha = 0$, and in the stratosphere

$$\varpi \propto e^{\lambda z}. \quad (23)$$

The exponent, λz , must be real for the energy to be finite. If it is unreal then it is readily shown that energy is transmitted upwards or downwards in the stratosphere; upwards transmission only can occur since no means for reflexion is assumed to exist above the tropopause, and if energy travels upwards then the lower sign is again chosen, but then no oscillation could be observed on the ground at a great distance from the source.

Free waves

If the oscillations take place over horizontal stationary ground then the boundary condition is

$$w_1 = 0, \quad \text{and it follows from equation (14) that } \varpi'_1 = 0. \quad (24)$$

Equations (23) and (24) impose two boundary conditions upon the solution of (16) in the troposphere, for at the tropopause p and w are continuous, and so ϖ and $\varpi'/(g\tau'_0/\tau_0 + \sigma^2)$ are continuous. But by equation (23), at the tropopause on the stratosphere side $\varpi' = \lambda\varpi$, therefore bearing in mind that τ'_0/τ_0 is discontinuous at the tropopause and expressing this condition in terms of χ , the boundary condition at $z = z_2$ imposed upon equation (16) is

$$\chi'_2 = \left\{ \frac{m}{2z_2} - \frac{1}{2(z_2 + b)} + \lambda \frac{\sigma^2 c_s^2 - (1 - \mu)(\gamma - 1)g}{\sigma^2 c_s^2 - (\gamma - 1)g} \right\} \chi_2. \quad (25)$$

The condition on χ at $z = z_1$, derived from (24), is

$$\chi'_1 = \left\{ \frac{m}{2z_1} - \frac{1}{2(z_1 + b)} \right\} \chi_1. \quad (26)$$

If σ is given, then there exists a unique value of k , denoted by k^* , for which both boundary conditions are satisfied. Since the solution to (16) cannot be expressed in finite terms or in terms of tabulated functions the equation was integrated numerically from $z = z_2$, the initial conditions being made to satisfy (25), making two or more trial values of k for a given σ , and obtaining by interpolation the value of k which would satisfy (26). After some preliminary exploration it was found possible to obtain k^* with great accuracy from only two trial values.

There is, however, one disadvantage of this method of direct integration in the particular case of equation (16), namely, that the coefficient S becomes infinite when $z = -b$. This happens within the range of integration for a small fraction of the range of σ under investigation, and numerical integration is impossible in and near to this fraction. It was therefore necessary to use Pekeris's equation (1948, p. 147, eqn. (9)) for the hydrodynamical divergence for one of the values of σ , and for two other values the values of k^* already found were checked. Pekeris's equation (9) when transformed to be suitable for numerical integration is

$$\psi'' = \left[\frac{A^2 - A}{z^2} - \left(B_1 + B_2 \frac{k^2}{\sigma^2} \right) \frac{1}{z} + k^2 \right] \psi,$$

where

$$A = \frac{1}{2} \left(1 + \frac{\gamma}{(\gamma - 1)\mu} \right),$$

$$B_1 = -\sigma^2/(\gamma - 1)\mu g,$$

$$B_2 = -(1 - \mu)g/\mu,$$

$$\psi = z^A \chi,$$

and χ is now the divergence of velocity defined by Pekeris. The boundary conditions are

$$\psi'_2 = \left(\nu + \frac{A}{z_2} \right) \psi_2,$$

$$\psi'_1 = \left\{ -g \left(\frac{k^2}{\sigma^2} - \frac{z_2 \gamma}{z_1 c_s^2} \right) + \frac{A}{z_1} \right\} \psi_1,$$

$$\nu = \lambda + \frac{(\gamma - 1)g}{c_s^2},$$

λ being already defined by equation (21).

The integration is slightly simpler numerically using this equation, but since the pressure is a function of the divergence and its vertical derivative, the results require more calculation to give the pulse. There is altogether nothing to choose between the use of Pekeris's equation and equation (12).

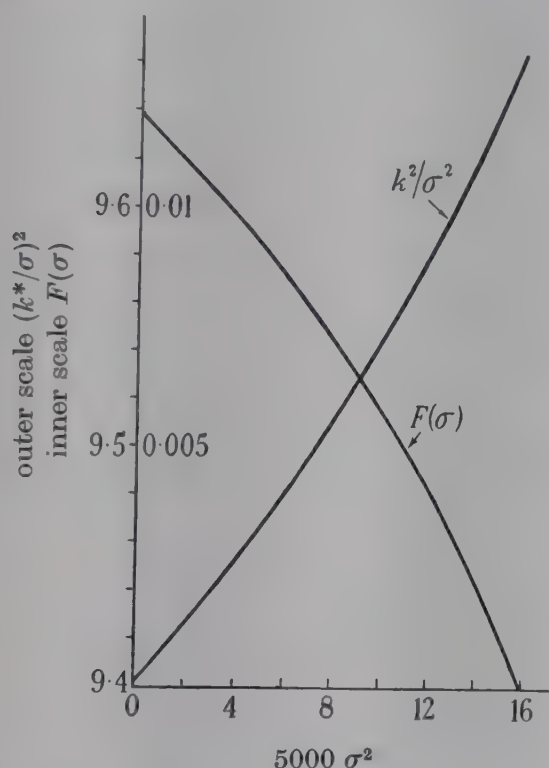


FIGURE 2. $(k^*/\sigma)^2$ and $F(\sigma)$ as functions of σ^2 .

For convenience $(k^*/\sigma)^2$ was found as a function of σ^2 and the results are given in figure 2 and table 1, for the atmosphere with the following constants:

$$g = 980.6 \text{ cm. sec.}^{-2}; \quad T_2 = T_s = 229.53^\circ \text{ K}; \quad T_1 = 286.91^\circ \text{ K};$$

$$z_2 - z_1 = 9.6137 \text{ km.}; \quad \gamma = 1.403.$$

In the numerical work the unit of length was taken as 0.9806/1.02 km.; when the depth of the troposphere was 10 units, $g = 1.02$, $c_s^2 = 10^{-1}$, and $\frac{1-\mu}{\mu} = 0.64424$.

In finding k^* as a function of σ for the free waves, that is, waves propagated over level ground, no reference is made to the form of equation (11). If in the place of a

spherical earth the ground had been taken as an infinite plane and the motion assumed to be independent of one horizontal co-ordinate y , then (11) takes the form

$$\frac{\partial^2 X}{\partial x^2} + k^2 X = 0, \tag{27}$$

where $\varpi \equiv X(x) \cdot \varpi(z)$,

and (12) is also satisfied. In this case the horizontal wave-length of the oscillations is $2\pi/k$, and σ/k^* is the horizontal velocity of the free waves. The meaning of k in equation (12) becomes clear only in terms of the solution of equation (11) and the way in which the explosive source is defined in terms of that solution.

TABLE 1

The unit of length for the 2nd and 3rd columns is $\frac{0.9806}{1.02}$ km. Three more significant figures were obtained for k^*/σ than are given in the table. The cut-off group velocity was obtained by extrapolation.

5000 σ^2	free waves		relative intensity $F(\sigma) \times 10^3$
	phase veloc. ⁻¹ k^*/σ	group veloc. ⁻¹ $dk^*/d\sigma$	
0	3.0663	3.0663	11.918
1	3.0682	3.0721	11.410
2	3.0702	3.0783	10.897
3	3.0723	3.0849	10.363
4	3.0744	3.0919	9.803
5	3.0766	3.0994	9.200
6	3.0790	3.1074	8.611
7	3.0814	3.1159	7.973
8	3.0839	3.1251	7.277
10	3.0893	3.1457	5.786
12	3.0952	3.1699	4.103
14	3.1018	3.1988	2.173
15.940	3.1089	3.231	0

Cut-off frequency

The k, σ plane is divided into two regions in which the condition

$$\left(\frac{(2-\gamma)g}{2c_s^2}\right)^2 + \left(\frac{(\gamma-1)g}{c_s^2} - \sigma^2\right)\left(\frac{1}{c_s^2} - \frac{k^2}{\sigma^2}\right) > 0 \tag{28}$$

is or is not satisfied. We are concerned only with the region in which it is satisfied for, as explained above, the oscillations must possess finite kinetic energy. The curve $k = k^*(\sigma)$ is found to lie partly in each region. For long waves ($\sigma = 0$) it is known from the work of Taylor and others that there is only one value of k^* , there being for this model atmosphere only one mode of oscillation. The present numerical exploration confirms that of Pekeris, that for each σ there is only one value of k^* and that there is a 'cut-off' frequency σ_c such that only when $\sigma < \sigma_c$ is (28) satisfied when $k = k^*$.

The physical meaning of this is that if they are excited by a source on the ground, oscillations of frequency less than σ_c are reflected down at the tropopause, while the

energy of those of greater frequency is transmitted indefinitely upwards in the stratosphere. It is in this property that our model atmosphere differs from the actual atmosphere which is known to have warmer layers at heights above 30 km. capable of trapping oscillations of audible frequencies. The results derived from the long-wave end of the spectrum only will bear good resemblance to observations.

It is found that the velocity σ/k^* of the free waves in our model atmosphere is equal to the velocity of sound c at some level in the troposphere depending upon σ . Their speed is slower than that of sound at the ground but faster than that in the stratosphere. The actual period at the cut-off frequency is found to be

$$2\pi/\sigma_c = 111.28 \text{ sec.}$$

4. THE DISPERSION OF A PULSE

It now remains only to impose a boundary condition at the ground which will correspond to an explosion. The pole ($r = 0, \theta = 0$) of the co-ordinates is taken as the centre of the explosion, which is represented by an upward movement of the ground over a finite (or infinitesimal) region for a finite (or infinitesimal) time so that a finite calculable volume of ground is introduced into the atmosphere. This corresponds to the sudden creation of gas by high explosive, the emission of gas or steam by a volcano, the sudden introduction into the atmosphere of a meteorite and its aura, or to the sudden generation of heat which causes a large and rapid expansion. The essential characteristic of an explosion for the present purpose is the rapid introduction of a volume into the atmosphere, and it will be seen that provided it is rapid and local enough, the pulse observed at a great distance depends very little upon the exact time taken to introduce it or the horizontal extent of the source.

We make use of the Fourier integral

$$w_1 = B \int_0^\infty d\sigma \int_0^\infty dk \Sigma(\sigma, t) K(k, r) e^{ikr} (e^{i\sigma t} + e^{-i\sigma t}),$$

(real part only) (29)

which, by a suitable choice of the functions Σ and K gives to w_1 a finite value near to $r = 0, t = 0$, and a negligible value elsewhere. The boundary condition (29) which defines the vertical velocity of the air at the ground thus represents a rapid, local elevation of the ground introducing a known volume into the atmosphere. B is a constant used to determine the magnitude.

Before inserting specific functions Σ and K we derive the formula for the pressure pulse p_1 observed at the ground at a great distance (large r). If

$$w = W(z), \quad \varpi = \Phi(z) \tag{30}$$

are corresponding solutions of equations (12) and (14) which satisfy the upper-boundary condition, then the vertical velocity at any height, with equation (28) as the lower-boundary condition, is

$$w = B \int_0^\infty d\sigma \int_0^\infty dk \Sigma K e^{ikr} (e^{i\sigma t} + e^{-i\sigma t}) \frac{W}{W_1},$$

so that by equation (30)

$$\varpi = B \int_0^\infty d\sigma \int_0^\infty dk \Sigma K e^{ikr} (e^{i\sigma t} + e^{-i\sigma t}) \frac{\Phi}{W_1}. \quad (31)$$

For free waves $k = k^*(\sigma)$, $W_1 = 0$, $\Phi \neq 0$, so that the integration with respect to k may be performed by the method of residues, k^* being the pole of the integrand. The disturbance of pressure at the ground ($z = z_1$) is obtained from (31) and (1), giving

$$p_1 = B \frac{\rho_{01}}{\tau_{01}} \int_0^\infty d\sigma \Sigma (e^{i\sigma t} + e^{-i\sigma t}) \left\{ \int_\Gamma dk K \frac{\Phi_1}{W_1} e^{ikr} + \pi i e^{ik^*r} K^* / \left[\frac{\partial W_1}{\partial k \Phi_1} \right]_{k=k^*} \right\}, \quad (32)$$

Γ being a contour in the upper half k -plane from the origin to $+\infty$ but not coinciding with the real axis except at its end-points. The first term in the bracket, though possibly appreciable near the source ($r = 0$) becomes negligible for large r on account of the oscillatory nature of e^{ikr} and negative real part of ikr , while $K(\Phi_1/W_1)$ is a slowly varying function of k for all r . For values of σ which do not give a real value of k^* , i.e. for $\sigma > \sigma_c$, the value of k^* has not been found, but for large r the factor e^{ik^*r} renders the contribution to p_1 negligible. The first term, together with the second term for $\sigma > \sigma_c$, represents a pulse which is perceptible at the ground near the source only, and includes the component oscillations of all those frequencies which are not trapped at the tropopause. The pulse observed on the ground at a great distance is therefore

$$p_1 = \pi i B \frac{\rho_{01}}{\tau_{01}} \int_0^{\sigma_c} d\sigma (e^{i\sigma t} + e^{-i\sigma t}) e^{ik^*r} \Sigma K / \left[\frac{\partial W_1}{\partial k \Phi_1} \right]_{k=k^*}. \quad (33)$$

Now by equations (11) and (30)

$$\begin{aligned} \left[\frac{\partial W_1}{\partial k \Phi_1} \right]_{k=k^*} &= \left[\frac{\partial}{\partial k} \frac{i\sigma \Phi_1'}{(g\tau_{01}' + \sigma^2 \tau_{01}) \Phi_1} \right]_{k=k^*} \\ &= \frac{i\sigma}{g\tau_{01}' + \sigma^2 \tau_{01}} \left[\frac{\partial}{\partial k} \left(\frac{\chi_1'}{\chi_1} \right) \right]_{k=k^*} \quad \text{by (15)} \\ &= \frac{4i}{\tau_{01}} / F(\sigma), \end{aligned} \quad (34)$$

where

$$2/F(\sigma) = \frac{k/\sigma}{g\tau_{01}'/\tau_{01} + \sigma^2} \left[\frac{\partial}{\partial(k^2/\sigma^2)} \left(\frac{\chi_1'}{\chi_1} \right) \right]_{k=k^*}. \quad (35)$$

$F(\sigma)$ is the relative intensity in the pulse of the component oscillations of frequency σ , in the case $\Sigma = K = 1$. The quantity in the square bracket is evaluated numerically in the course of finding k^* by making trial values of k^2/σ^2 for a given σ^2 , and $F(\sigma)$ is given in figure 2 and table 1 as a function of σ^2 . The numerical integration of equation (16) was performed from the tropopause downwards because it is required in (35) that χ_1 shall arise from a function Φ , given by (30), satisfying the upper-boundary condition. We now have

$$p_1 = \frac{1}{4} \pi \rho_{01} B \int_0^{\sigma_c} F \Sigma K^* e^{ik^*r} (e^{i\sigma t} + e^{-i\sigma t}) d\sigma, \quad (36)$$

which gives p_1 as a function of r and t .

The succeeding section discusses some particular forms of Σ and K .

5. THE FORM OF THE EXPLOSIVE SOURCE

The region over which $w_1 \neq 0$ is only a small fraction of the earth's surface, and can be considered flat; so that analytically the source has the same form as if it were on an infinite plane earth. In such circumstances if the oscillations were independent of the horizontal co-ordinate y , an instantaneous line source such that w_1 is zero at all points except $x = 0$ and at all times except $t = 0$, but is integrably infinite at $x = 0, t = 0$, is represented by equation (29) with

$$\Sigma = K = 1, \quad r = x, \quad (37)$$

giving
$$w_1 = 2B \int_0^\infty \cos \sigma t d\sigma \int_0^\infty \cos kx dk, \quad (38)$$

$$p_1 = \frac{1}{4}\pi\rho_{01} B \int_0^{\sigma_c} F(\sigma) e^{ik^*x} (e^{i\sigma t} + e^{-i\sigma t}) d\sigma \quad (\text{real part}). \quad (39)$$

It may be noted that $\cos kx$ is a solution of (26), $2\pi/k$ is the horizontal wave-length and σ/k the horizontal phase velocity of the component oscillations.

When r is small equation (11) becomes Bessel's equation of zero order, and an instantaneous point source at $r = 0, t = 0$ is represented by the well-known integral (used by Pekeris)

$$w_1 = 2B \int_0^\infty \cos \sigma t d\sigma \int_0^\infty J_0(kr) k dk. \quad (40)$$

The volume introduced is

$$\int_0^\infty dt \int_0^\infty 2\pi r w_1 dr = 8\pi^2 B$$

(Lamb 1932, art. 102) and

$$p_1 = \frac{1}{4}\pi\rho_{01} B \int_0^{\sigma_c} F(\sigma) J_0(k^*r) k^* (e^{i\sigma t} + e^{-i\sigma t}) d\sigma. \quad (41)$$

Here $K e^{ikr}$ in (29) has been replaced by $J_0(kr) k$, and the argument used to dispose of the first term in (33) is valid because the Bessel function oscillates like the cosine. When r is sufficiently large, either the integrand is negligible because k^* is small when σ is near to 0 or else $J_0(k^*r)$ may be replaced by its asymptotic form, giving

$$p_1 = \frac{1}{4}\pi\rho_{01} B \int_0^{\sigma_c} F(\sigma) \left(\frac{2k^*}{\pi r}\right)^{\frac{1}{2}} e^{i(k^*r - \frac{1}{2}\pi)} (e^{i\sigma t} + e^{-i\sigma t}) d\sigma. \quad (42)$$

However, when r is large, Bessel functions are no longer solutions of equation (11) and it is therefore necessary to replace $r^{-\frac{1}{2}}$ by $(a \sin \theta)^{-\frac{1}{2}}$, thus substituting the asymptotic form of the solution of equation (11) (Jeffreys & Jeffreys 1946, art. 24.15) for the asymptotic form of the Bessel function. The factor $k^{\frac{1}{2}}$ means that as compared with a line source, a point source favours the shorter wave-lengths. The form taken by (42) on a spherical earth is obtained by putting into the expression (36) the values

$$\Sigma = 1, \quad K = \left(\frac{2k}{a \sin \theta}\right)^{\frac{1}{2}} e^{-\frac{1}{2}i\pi}. \quad (43)$$

This is the form of source for which the pulse form at various distances has been calculated. The modification required when the source is diffused in space or time is

illustrated by three examples, in each of which the relative intensity of that part of the pulse due to the longer waves is increased. It is reasonable to suppose that this would be the effect of most kinds of spreading of the source in space or time. Since the longer waves travel faster the effect is simply to intensify the earlier part of the pulse at the expense of the later part.

(a) A source of intensity gradually increasing to a maximum at $t = 0$ and then dying away, the strength at time t being proportional to $U/(U^2 + t^2)$, which is the real part of $\int_0^\infty e^{-\sigma U + i\sigma t} d\sigma$, is obtained by putting $\Sigma = e^{-\sigma U}$, where U is a constant, being the time taken for the source to decrease from its maximum to half its maximum strength.

(b) If the intensity is proportional to $\cos st$ in $-\pi/2s < t < \pi/2s$ and is zero at all other times, we write

$$\Sigma = \frac{2s}{s^2 - \sigma^2} \cos \frac{\pi\sigma}{2s}. \quad (44)$$

This again is a decreasing function of σ when σ is positive.

(c) If instead of a point source the intensity is uniform inside $r = b$ and zero outside we find (cf. Lamb 1932, art. 102) that it is necessary to write

$$K = \frac{2}{\pi b} J_1(kb) \left(\frac{2}{k\pi a \sin \theta} \right)^{\frac{1}{2}} e^{-\frac{1}{2}i\pi} \quad (45)$$

which multiplies the factor given in (43) by $2J_1(kb)/kb$. The relative intensity of the long waves is again increased.

In these instances the constants U , s and b must be chosen so that Σ and K^* are slowly varying functions of σ , as is $F(\sigma)$, for upon that fact depends the method of evaluating the pulse numerically.

6. EVALUATION OF THE INTEGRAL $I = \int_0^{\sigma_0} F \Sigma K e^{ikx} (e^{i\sigma t} + e^{-i\sigma t}) d\sigma$ (EQUATION (36))

It is necessary to describe only briefly how the method given elsewhere (Scorer 1950) was used to evaluate the integral (36).

Henceforth we write k for k^* , K for K^* , and k' for $dk^*/d\sigma$.

It became evident as the values of k^2/σ^2 were found, and plotted against σ^2 , that as $\sigma \rightarrow 0$ then $k'' \rightarrow 0$ also, and application of the principle of stationary phase, suggested by Pekeris for this problem, could not be made in the usual way,† for it is only justifiable if

$$|k'''| \ll |k''|^{\frac{1}{2}}. \quad (46)$$

But since $k^{1/2} \rightarrow 0$ as $\sigma \rightarrow 0$, the method is adequate if $kx - \sigma t$ is expressed as a cubic in σ instead of the usual quadratic.

Thus if $k(\sigma_0) = k_0$ is the solution of

$$k'x - t = 0 \quad (47)$$

† As given, for instance, in Lamb (1932, art. 241).

($k'x + t = 0$ has in the present case no solution, so that the term in $e^{i\sigma t}$ has no point of stationary phase and is therefore neglected), and

$$k = k_0 + (\sigma - \sigma_0) k'_0 + \frac{1}{2}(\sigma - \sigma_0)^2 k''_0 + \frac{1}{6}(\sigma - \sigma_0)^3 k'''_0, \quad (48)$$

$$\text{then} \quad kx - \sigma t = a + b\sigma + c\sigma^2 + d\sigma^3, \quad (49)$$

$$\text{where} \quad \left. \begin{aligned} a &= (k_0 - \sigma_0 k'_0 + \frac{1}{2}\sigma_0^2 k''_0 - \frac{1}{6}\sigma_0^3 k'''_0) x, \\ b &= (-\sigma_0 k''_0 + \frac{1}{2}\sigma_0^2 k'''_0) x, \\ c &= (\frac{1}{2}k''_0 - \frac{1}{2}\sigma_0 k'''_0) x, \\ d &= \frac{1}{6}k'''_0 x. \end{aligned} \right\} \quad (50)$$

Thus, given x , by (47), σ_0 is a function of t , and following the method given by Scorer (1950) we obtain

$$I = \frac{\pi}{l} F_0 \Sigma_0 K_0 \exp \left\{ i \left(a - \frac{bc}{3d} \right) \right\} L(z), \quad (51)$$

which gives I as a function of σ_0 and therefore of t for a given x , where

$$\left. \begin{aligned} l &= (3d)^{\frac{1}{3}}, \\ z &= (3d)^{-\frac{1}{3}} (b - c^2/3d), \end{aligned} \right\} \quad (52)$$

$$\text{and} \quad L(z) = \text{Ai}(z) + i \text{Gi}(z)$$

$$= \frac{1}{\pi} \int_0^\infty \exp \{ i(uz + \frac{1}{3}u^3) \} du. \quad (53)$$

Tables of $\text{Ai}(z)$ are given by Miller (1946) and of $\text{Gi}(z)$ by Scorer (1950).

To evaluate the derivatives of k most accurately by numerical methods we write

$$k^2/\sigma^2 = V(\sigma^2), \quad V^{(n)} = \frac{d^n V}{d(\sigma^2)^n}, \quad (54)$$

and obtain

$$\left. \begin{aligned} k/\sigma &= V^{\frac{1}{2}}, \\ k &= \sigma V^{\frac{1}{2}}, \\ k' &= V^{\frac{1}{2}} + \sigma^2 V^{-\frac{1}{2}} V', \\ k'' &= \sigma V^{-\frac{1}{2}} \{ 3V' - \sigma^2 V^{-1} V'^2 + 2\sigma^2 V'' \}, \\ k''' &= V^{-\frac{1}{2}} \{ 3V' + 12\sigma^2 V'' + 4\sigma^4 V''' - 6\sigma^2 V^{-1} (V'^2 + \sigma^2 V' V'') + 3\sigma^4 V^{-2} V'^3 \}. \end{aligned} \right\} \quad (55)$$

V' , V'' , V''' were obtained by numerical differentiation of $V(\sigma^2)$. This procedure, though most undesirable from the computation point of view, was unavoidable, but the higher derivatives of V only made small contributions in the formulae (55). This was the main reason for calculating k^2/σ^2 at equal intervals of σ^2 , for thereby a very smooth function is obtained.

Equations (47) and (51) were then used to obtain I and therefore p_1 , as a function of t for certain chosen values of r .

7. COMPARISON OF RESULTS WITH OBSERVATION

The pulses to be observed at five different great circle distances from the explosion are given in figures 3 to 7. The units of pressure are microbars (dyne cm.⁻²) per cu.km. of explosion. If all the energy to cause the expansion were supplied as heat the amount per cu.km. would be approximately 3.8×10^{21} ergs, independent of the volume to which the heat is applied, assuming that the heating took place at constant pressure. If the volume is introduced mechanically the energy required depends very much upon the rate of introduction, but it is probably of the same order of magnitude as if the energy is supplied as heat if the expansion is rapid enough to be termed an explosion. Details of these calculated pulses are given in table 2, the time zero in each case being placed at the time of arrival of long waves. The time of cut-off is the time of

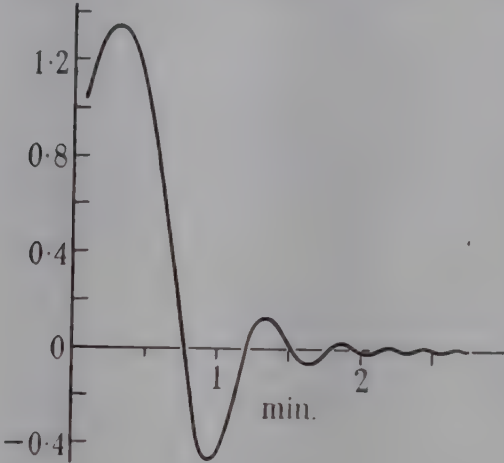


FIGURE 3

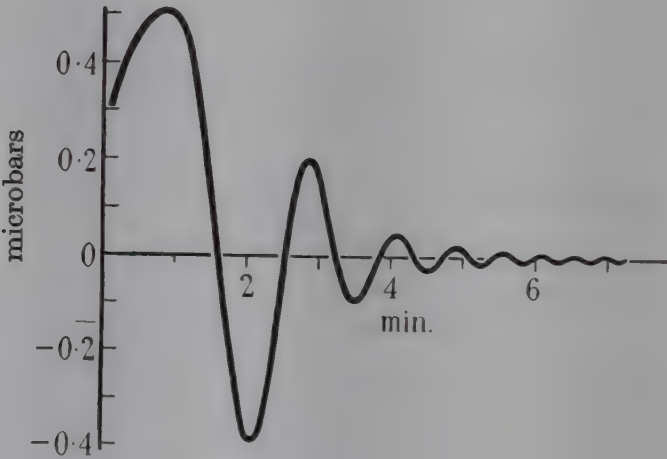


FIGURE 4

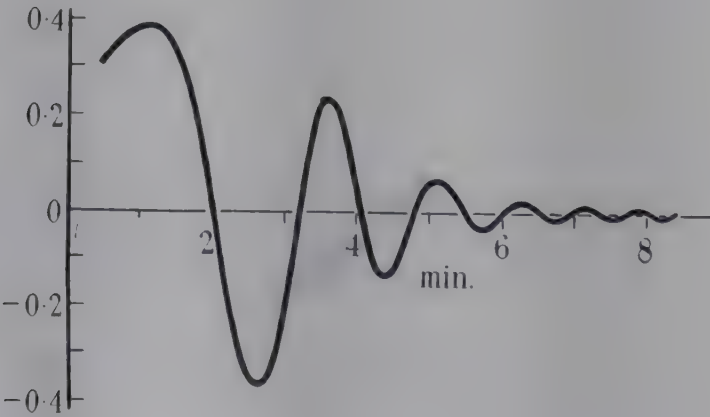


FIGURE 5

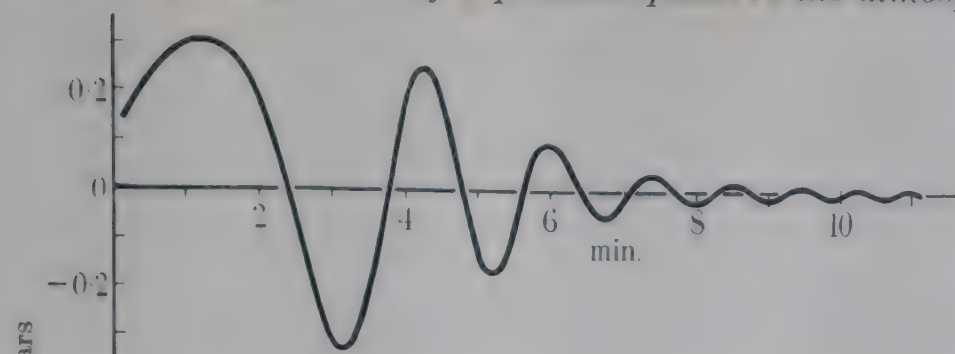


FIGURE 6

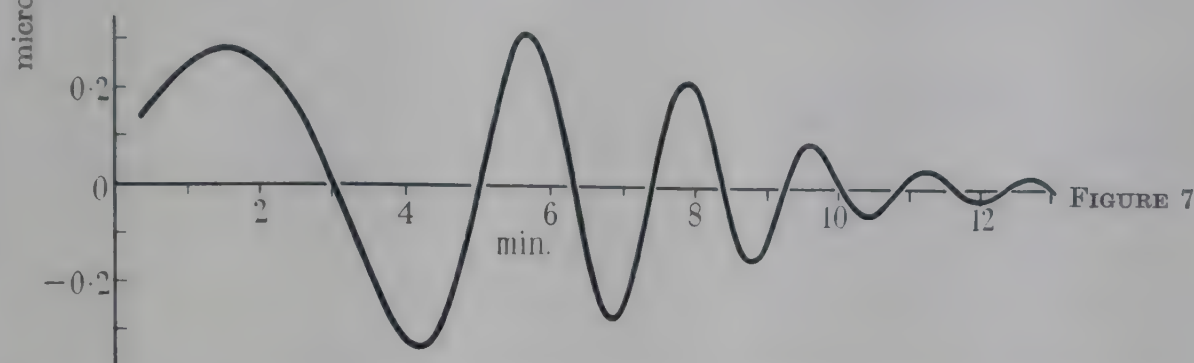


FIGURE 7

FIGURES 3 to 7. The pulses at five different great circle distances from the explosion.

arrival of waves travelling with the group velocity corresponding to the cut-off frequency. The amplitude of the first five half-oscillations is the difference in pressure between successive stationary values. Only in figure 3 is the complete pulse form given; in the other cases the oscillations continue with decreasing amplitude and increasing frequency to a period of 12.7 sec. at the cut-off time. The form of the pulse before the first crest is uncertain for the stationary phase method is inadequate near $\sigma_0 = 0$ because the factor $k^{\frac{1}{2}}$ in the integrand is rapidly varying near $k = 0$; but in any event, an instrument recording the pulse would be fitted with some form of leakage so that slow changes in pressure are not shown. Thus the first crest is probably too low on the microbarograph traces for the English stations shown with fourfold magnification in figures 8 to 11. The rise in pressure from the first trough to the second crest has been taken as well as the fall from the first crest to the first trough when theory and observation are compared to give an estimate of the volume of the explosion of the Siberian meteorite, except in the case of stations around 970 km. distant when the fall from the first crest to the first trough only is taken. In table 3 are briefly set out the characteristics of the best available observations. The Siberian stations were twelve in number ranging from 660 to 1230 km. distant. The average distance was 990 km. The pressure-pulse amplitude varied from 0.40 to 2.45 and was not well related to the distance. This information is given in the paper by Astapowitsch (1934), whence also figure 12 was taken, showing the pulse observed near Leningrad.

There appears to be a loss of well over half the energy of the pulse between the Siberian and English stations. This may be attributed to various causes besides mechanical dissipation; the prevailing wind and temperature regime may cause refraction and distortion of the wave fronts; and the siting of the station relative to these and to mountain ranges may influence the pulse received. An expansion of 1000 cu.km. would be caused by 3.8×10^{24} ergs if all the energy were supplied as heat, whereas no previous estimate of the energy exceeds 5×10^{21} ergs.

In this estimate we have calculated the energy required to be supplied as heat in order to produce the observed pulse, but only a fraction of this energy passes into the kinetic energy of the pulse. The energy of all those component frequencies for which $\sigma > \sigma_c$ escapes into the upper stratosphere and is dissipated by viscosity. Work must be done merely to introduce the volume: approximately 10^{24} ergs would be required to introduce 1000 cu.km. so slowly that no pulse were created at all. Previous estimates, summarized by Astapowitsch, of the energy depend on calculating the energy of the air wave or the work done in felling the observed number of trees, that is, the

FIGURE 8.
Meteorological
Office.

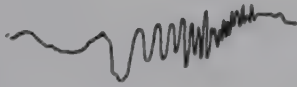
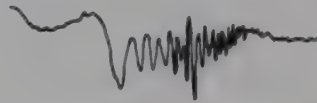


FIGURE 9.
Reading.



30 min.

FIGURE 10.
Petersfield.

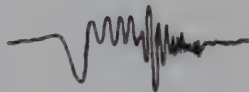


FIGURE 11.
South
Kensington.

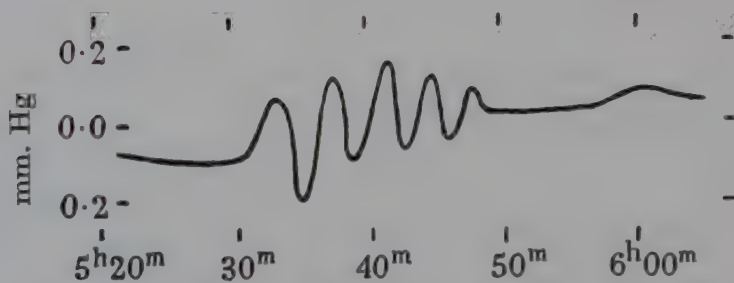
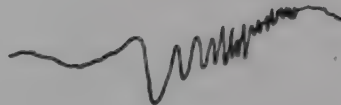


FIGURE 12. Diagram of record of air waves caused by the explosion on the occasion of the fall of the meteorite on 30 June 1908 according to the records of the Sprung microbarograph at the Slutsk Geophysical Observatory near Leningrad.

energy required for one of the many phenomena produced, but no estimate was made of what proportion of the total energy was expended upon it. Alternatively, they depend upon comparison with the Krakatoa eruption for which no more reliable estimate exists, or with earthquakes. We are concerned here with the total amount of energy communicated to the atmosphere; and since the energy of the earthquake shock must be added to this it appears that the great magnitude of the phenomenon has not hitherto been appreciated.

The time interval between successive nodes is not in good agreement with observation. The atmosphere on the occasion seems to have been more dispersive than the model chosen, but even so no explanation is offered as to why the oscillations at Leningrad should have been slower than those observed in England.

Figure 13 shows, in an enlarged scale, the form of the pulse observed in England. It was derived by Whipple (1930) from six microbarograms, four of which are given

TABLE 2. CHARACTERISTICS OF THE CALCULATED PULSES

figure	great circle distance in km.	interval between explosion and time zero		time of cut-off min. sec.	amplitude of successive half-oscillations (microbars)		time interval between successive nodes (sec.)		no. of complete oscillations before cut-off
		min. sec.	min. sec.						
3	961	51	6	2 45	1.80, 0.59, 0.19, 0.08, 0.03		66, 26, 19, 15, 12, 10		5
4	3650	194	1	10 29	0.88, 0.57, 0.29, 0.14, 0.08		98, 56, 41, 36, 29, 25		19
5	5607	298	2	16 7	0.74, 0.59, 0.36, 0.19, 0.10		122, 71, 51, 46, 39, 33		30
6	7691	408	50	22 6	0.64, 0.58, 0.42, 0.26, 0.15		144, 83, 61, 49, 47, 41		42
7	13290	706	28	38 11	0.62, 0.66, 0.60, 0.50, 0.37		177, 124, 80, 65, 56, 51		70

TABLE 3. CHARACTERISTICS OF THE PULSES OBSERVED ON THE OCCASION OF THE FALL OF THE
GREAT SIBERIAN METEORITE IN 1908

station	distance (km.)	time to first trough (min.)	amplitude of successive half-oscillations	Time interval between successive nodes (min.)		volume of explosion derived from	
						1st half oscillation (cu.km.)	2nd half oscillation (cu.km.)
Siberian	(average) 990	—	1.15 mm. mercury	—		880	—
Leningrad	3740	195.2	0.275, 0.275, 0.25, 0.21, 0.21 mm.	2.0, 2.0, 2.0, 1.9, 1.8, 1.6		430	660
(Slutsk)			mercury				
English	(average) 5752	(average) 301	197, 214, 101, 73, 62 microbars	2.5, 1.3, 1.0, 1.0, 1.0, 1.0		270	360

in figures 8 to 11. The sudden and more rapid oscillations which are completely absent from the calculated curves may reasonably be supposed to be due to the warm layers above 30 km., which have not been taken into account. Whipple remarks that these waves travelled considerably faster than the sound waves known on occasions to have been reflected by these layers, but then only a single reflexion took place, and the height of reflexion was a considerable fraction of the distance between the explosion and the observer; the propagation was thought of in terms of rays of sound which were inclined to the ground. The oscillations which are observed at a great distance may not legitimately be studied by geometrical optics which is suitable for high-frequency oscillations only; gravity now becomes all important, and the wave-fronts are vertical; the path of a wave-front element from the region of the explosion to the observer is horizontal, and it is not surprising to find a greater horizontal component to the velocity of propagation than in the case of a single reflexion at high levels. The labour involved in the numerical work necessary to study this has so far been prohibitive. All that can be concluded on the basis of the present work is that this sudden, second, set of oscillations was probably not due to a separate explosion because at such a great distance the pulses for two separate explosions would be identical in form.

The kinetic-energy density of the free waves dies away exponentially upwards in an isothermal stratosphere when $\sigma < \sigma_c$, but dies away more slowly as σ increases until at $\sigma = \sigma_c$ it is constant at all heights; $F(\sigma)$ therefore decreases to zero as σ increases to σ_c , for at $\sigma = \sigma_c$ the energy of the oscillations is infinite unless the amplitude is zero. If the warmer layers at higher levels were taken into account then the cut-off would not occur until a higher value of σ were reached; therefore $F(\sigma)$ would assume a greater value for those frequencies which make up the later part of the calculated pulse, and the amplitude of the oscillations would die away less rapidly. This may, in part, explain the difference in form between the calculated and observed pulses. No such ready argument has been found to deduce the effect of the warm upper layers of the stratosphere on the period of the oscillations in the later part of the pulse.

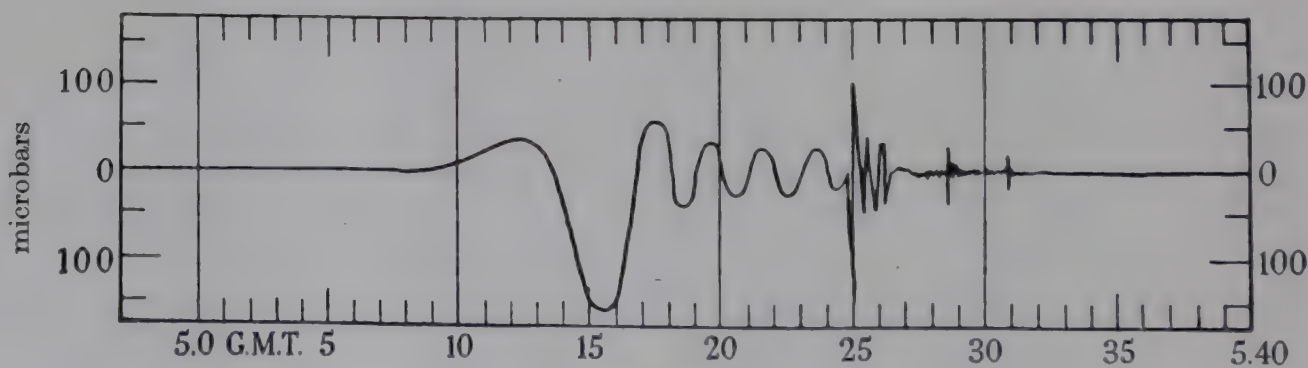


FIGURE 13.

Comparison with the Krakatoa air wave is difficult because there was probably not a single sudden explosion, nor were instruments set up capable of displaying the more rapid oscillations. There was undoubtedly one explosion greater than the others, but the others were great enough to produce observable pulses. The interval between the

explosions was of the same order of magnitude as the time taken for one pulse to pass a distant point, so that the separate pulses interfered. The best observations were, furthermore, made at stations of very different latitude from each other and from the eruption, so that the pulse form would be distorted by passage through regions with large variations in wind and temperature structure.

Finally, two remarks may be made concerning the calculated pulse form. The local period of oscillation at any part of the pulse does not correspond to the period of the waves, travelling with the group velocity and arriving at that instant. The cut-off frequency has a period of 111.28 sec. which is nearly ten times the period of the oscillations at the termination of the pulse. Pekeris found a cut-off period of about 2 min. for a slightly different model atmosphere.

It has been assumed that propagation takes place without loss of energy, and the amplitude at 7691 km. (figure 6) is smaller than at 13290 km. (figure 7) because the wave front in the former instance is extended round an equatorial belt (the explosion being taken as pole). It contracts subsequently to a small zone of higher latitude and the amplitude increases.

The author wishes to thank Professor Sir Geoffrey Taylor, F.R.S., for fundamental discussions on the problem, Dr M. V. Wilkes for advice on the numerical work, and Miss C. M. Munford for assistance with the computation. The Royal Meteorological Society has kindly given permission for the reproduction of figures 12 and 13; and the author is indebted to the Director of the Meteorological Office for the loan of original microbarograms and to the Controller of His Majesty's Stationery Office for permission to publish enlarged copies of some in figures 8 to 11. These microbarograms are Crown copyright. Thanks are also due to the Ministry of Education for a Further Education and Training grant which has enabled this study to be made.

REFERENCES

- Astapowitsch, I. S. 1934 *Quart. J. R. Met. Soc.* **60**, 493.
 Jeffreys, H. & Jeffreys, B. 1946 *Mathematical Physics*. Cambridge University Press.
 Krakatoa Committee of the Royal Society 1888 *The Eruption of Krakatoa and subsequent phenomena*. London: Trübner and Co.
 Lamb, H. 1932 *Hydrodynamics*. Cambridge University Press.
 Miller, C. J. P. 1946 *B. A. Tables, Part Vol. B. The Airy Integral*.
 Pekeris, C. L. 1937 *Proc. Roy. Soc. A*, **158**, 650.
 Pekeris, C. L. 1939 *Proc. Roy. Soc. A*, **171**, 434.
 Pekeris, C. L. 1948 *Phys. Rev.* **73**, no. 2, 145.
 Scorer, R. S. 1949 *Quart. J. R. Met. Soc.* **75**, 41.
 Scorer, R. S. 1950 *Quart. J. Mech. App. Maths.* **3**, no. 1 (in Press).
 Taylor, G. I. 1929 *Proc. Roy. Soc. A*, **126**, 169 and 728.
 Taylor, G. I. 1936 *Proc. Roy. Soc. A*, **156**, 318.
 Whipple, F. J. W. 1930 *Quart. J. R. Met. Soc.* **56**, 287.
 Whipple, F. J. W. 1934 *Quart. J. R. Met. Soc.* **60**, 505.
 Whipple, F. J. W. 1935 *Quart. J. R. Met. Soc.* **61**, 285.
 Wilkes, M. V. & Weekes, K. 1947 *Proc. Roy. Soc. A*, **192**, 80.

The formation of a blast wave by a very intense explosion

I. Theoretical discussion

BY SIR GEOFFREY TAYLOR, F.R.S.

(Received 6 October 1949)

SUMMARY AND INTRODUCTION

This paper was written early in 1941 and circulated to the Civil Defence Research Committee of the Ministry of Home Security in June of that year. The present writer had been told that it might be possible to produce a bomb in which a very large amount of energy would be released by nuclear fission—the name atomic bomb had not then been used—and the work here described represents his first attempt to form an idea of what mechanical effects might be expected if such an explosion could occur. In the then common explosive bomb mechanical effects were produced by the sudden generation of a large amount of gas at a high temperature in a confined space. The practical question which required an answer was: Would similar effects be produced if energy could be released in a highly concentrated form unaccompanied by the generation of gas? This paper has now been declassified, and though it has been superseded by more complete calculations, it seems appropriate to publish it as it was first written, without alteration, except for the omission of a few lines, the addition of this summary, and a comparison with some more recent experimental work, so that the writings of later workers in this field may be appreciated.

An ideal problem is here discussed. A finite amount of energy is suddenly released in an infinitely concentrated form. The motion and pressure of the surrounding air is calculated. It is found that a spherical shock wave is propagated outwards whose radius R is related to the time t since the explosion started by the equation

$$R = S(\gamma) t^{\frac{2}{3}} E^{\frac{1}{3}} \rho_0^{-\frac{1}{3}},$$

where ρ_0 is the atmospheric density, E is the energy released and $S(\gamma)$ a calculated function of γ , the ratio of the specific heats of air.

The effect of the explosion is to force most of the air within the shock front into a thin shell just inside that front. As the front expands, the maximum pressure decreases till, at about 10 atm., the analysis ceases to be accurate. At 20 atm. 45 % of the energy has been degraded into heat which is not available for doing work and used up in expanding against atmospheric pressure. This leads to the prediction that an atomic bomb would be only half as efficient, as a blast-producer, as a high explosive releasing the same amount of energy.

In the ideal problem the maximum pressure is proportional to R^{-3} , and comparison with the measured pressures near high explosives, in the range of radii where the two might be expected to be comparable, shows that these conclusions are borne out by experiment.

SIMILARITY ASSUMPTION

The propagation and decay of a blast wave in air has been studied for the case when the maximum excess over atmospheric pressure does not exceed 2 atm. At great distances R from the explosion centre the pressure excess decays as in a sound wave proportionally to R^{-1} . At points nearer to the centre it decays more rapidly than R^{-1} . When the excess pressure is 0.5 atm., for instance, a logarithmic plot shows that it varies as $R^{-1.9}$. When the excess pressure is 1.5 atm. the decay is proportional to $R^{-2.8}$. It is difficult to analyze blast waves in air at points near the explosion centre because the initial shock wave raises the entropy of the air it traverses by an amount which depends on the intensity of the shock wave. The passage of a spherical shock wave, therefore, leaves the air in a state in which the entropy decreases radially so that after its passage, when the air has returned to atmospheric pressure, the air temperature decreases with increasing distance from the site of the explosion. For this reason the density is not a single-valued function of the pressure in a blast wave. After the passage of the blast wave, the relationship between pressure and density for any given particle of air is simply the adiabatic one corresponding with the entropy with which that particle was endowed by the shock wave during its passage past it. For this reason it is in general necessary to use a form of analysis in which the initial position of each particle is retained as one of the variables. This introduces great complexity and, in general, solutions can only be derived by using step-by-step numerical integration. On the other hand, the great simplicity which has been introduced into two analogous problems, namely, the spherical detonation wave (Taylor 1950) and the air wave surrounding a uniformly expanding sphere (Taylor 1946), by assuming that the disturbance is similar at all times, merely increasing its linear dimensions with increasing time from initiation, gives encouragement to an attempt to apply similar principles to the blast wave produced by a very intense explosion in a very small volume.

It is clear that the type of similarity which proved to be possible in the two above-mentioned problems cannot apply to a blast wave because in the latter case the intensity must decrease with increasing distance while the total energy remains constant. In the former the energy associated with the motion increased proportionally to the cube of the radius while the pressure and velocity at corresponding points was independent of time.

The appropriate similarity assumptions for an expanding blast wave of constant total energy are

$$\text{pressure, } p/p_0 = y = R^{-3}f_1, \quad (1)$$

$$\text{density, } \rho/\rho_0 = \psi, \quad (2)$$

$$\text{radial velocity, } u = R^{-\frac{1}{2}}\phi_1, \quad (3)$$

where R is the radius of the shock wave forming the outer edge of the disturbance, p_0 and ρ_0 are the pressure and density of the undisturbed atmosphere. If r is the radial co-ordinate, $\eta = r/R$ and f_1 , ϕ_1 and ψ are functions of η . It is found that these assumptions are consistent with the equations of motion and continuity and with the equation of state of a perfect gas.

The equation of motion is

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial r} = -\frac{p_0}{\rho} \frac{\partial y}{\partial r} \quad (4)$$

Substituting from (1), (2) and (3) in (4) and writing f'_1, ϕ'_1 for $\frac{\partial}{\partial \eta} f_1, \frac{\partial}{\partial \eta} \phi_1$,

$$-(\frac{3}{2}\phi_1 + \eta\phi'_1) R^{-\frac{1}{2}} \frac{dR}{dt} + R^{-4} \left(\phi_1 \phi'_1 + \frac{p_0}{\rho_0} \frac{f'_1}{\psi} \right) = 0. \quad (5)$$

This can be satisfied if

$$\frac{dR}{dt} = AR^{-\frac{1}{2}}, \quad (6)$$

where A is a constant, and

$$-A(\frac{3}{2}\phi_1 + \eta\phi'_1) + \phi_1 \phi'_1 + \frac{p_0}{\rho_0} \frac{f'_1}{\psi} = 0. \quad (7)$$

The equation of continuity is

$$\frac{\partial \rho}{\partial t} + u \frac{\partial \rho}{\partial r} + \rho \left(\frac{\partial u}{\partial r} + \frac{2u}{r} \right) = 0. \quad (8)$$

Substituting from (1), (2), (3) and (6), (8) becomes

$$-A\eta\psi' + \psi'\phi_1 + \psi \left(\phi'_1 + \frac{2}{\eta}\phi_1 \right) = 0. \quad (9)$$

The equation of state for a perfect gas is

$$\left(\frac{\partial}{\partial t} + u \frac{\partial}{\partial r} \right) (p\rho^{-\gamma}) = 0. \quad (10)$$

where γ is the ratio of specific heats.

Substituting from (1), (2), (3) and (6), (10) becomes

$$A(3f_1 + \eta f'_1) + \frac{rf'_1}{\psi} \psi' (-A\eta + \phi_1) - \phi_1 f'_1 = 0. \quad (11)$$

The equations (7), (9) and (11) may be reduced to a non-dimensional form by substituting

$$f = f_1 a^2 / A^2, \quad (12)$$

$$\phi = \phi_1 / A, \quad (13)$$

where a is the velocity of sound in air so that $a^2 = \gamma p_0 / \rho_0$. The resulting equations which contain only one parameter, namely, γ , are

$$\phi'(\eta - \phi) = \frac{1}{\gamma} \frac{f'}{\psi} - \frac{3}{2}\phi, \quad (7a)$$

$$\frac{\psi'}{\psi} = \frac{\phi' + 2\phi/\eta}{\eta - \phi}, \quad (9a)$$

$$3f + \eta f' + \frac{\gamma \psi'}{\psi} f(-\eta + \phi) - \phi f' = 0. \quad (11a)$$

Eliminating ψ' from (11a) by means of (7a) and (9a) the equation for calculating f' when f, ϕ, ψ and η are given is

$$f' \{ (\eta - \phi)^2 - f/\psi \} = f \{ -3\eta + \phi(3 + \frac{1}{2}\gamma) - 2\gamma\phi^2/\eta \}. \quad (14)$$

When f' has been found from (14), ϕ' can be calculated from (7a) and hence ψ' from (9a). Thus if for any value of η , f , ϕ and ψ are known their values can be computed step-by-step for other values of η .

SHOCK-WAVE CONDITIONS

The conditions at the shock wave $\eta = 1$ are given by the Rankine-Hugoniot relations which may be reduced to the form

$$\frac{\rho_1}{\rho_0} = \frac{\gamma - 1 + (\gamma + 1)y_1}{\gamma + 1 + (\gamma - 1)y_1}, \quad (15)$$

$$\frac{U^2}{a^2} = \frac{1}{2\gamma} \{\gamma - 1 + (\gamma + 1)y_1\}, \quad (16)$$

$$\frac{u_1}{U} = \frac{2(y_1 - 1)}{\gamma - 1 + (\gamma + 1)y_1}, \quad (17)$$

where ρ_1 , u_1 and y_1 represent the values of ρ , u and y immediately behind the shock wave and $U = dR/dT$ is the radial velocity of the shock wave.

These conditions cannot be satisfied consistently with the similarity assumptions represented by (1), (2) and (3). On the other hand, when y_1 is large so that the pressure is high compared with atmospheric pressure, (15), (16) and (17) assume the approximate asymptotic forms

$$\frac{\rho_1}{\rho_0} = \frac{\gamma + 1}{\gamma - 1}, \quad (15a)$$

$$\frac{U^2}{a^2} = \frac{2\gamma}{\gamma + 1} y_1, \quad (16a)$$

$$\frac{u_1}{U} = \frac{2}{\gamma + 1}. \quad (17a)$$

These approximate boundary conditions are consistent with (1), (2), (3) and (6); in fact (15a) yields, for the conditions at $\eta = 1$,

$$\psi = \frac{\gamma + 1}{\gamma - 1}, \quad (15b)$$

$$(16a) \text{ yields } f = \frac{2\gamma}{\gamma + 1}, \quad (16b)$$

$$\text{and } (17a) \text{ yields } \phi = \frac{2}{\gamma + 1}. \quad (17b)$$

ENERGY

The total energy E of the disturbance may be regarded as consisting of two parts, the kinetic energy

$$K.E. = 4\pi \int_0^R \frac{1}{2} \rho u^2 r^2 dr,$$

and the heat energy

$$H.E. = 4\pi \int_0^R \frac{pr^2}{\gamma - 1} dr.$$

In terms of the variables f, ϕ, ψ and η

$$E = 4\pi A^2 \left\{ \frac{1}{2} \rho_0 \int_0^1 \psi \phi^2 \eta^2 d\eta + \left(\frac{p_0}{a^2(\gamma-1)} \int_0^1 f \eta^2 d\eta \right) \right\},$$

or since $p_0 = a^2 \rho_0 / \gamma$, $E = B \rho_0 A^2$, where B is a function of γ only whose value is

$$B = 2\pi \int_0^1 \psi \phi^2 \eta^2 d\eta + \frac{4\pi}{\gamma(\gamma-1)} \int_0^1 f \eta^2 d\eta. \quad (18)$$

Since the two integrals in (18) are both functions of γ only it seems that for a given value of γ , A^2 is simply proportional to E/ρ_0 .

NUMERICAL SOLUTION FOR $\gamma = 1.4$

When $\gamma = 1.4$ the boundary values of f, ϕ and ψ at $\eta = 1$ are from (15*a*), (16*a*), (17*a*), $\frac{7}{8}$, $\frac{5}{8}$ and 6. Values of f, ϕ and ψ were calculated from $\eta = 1.0$ to $\eta_1 = 0.5$, using intervals of 0.02 in η . Starting each step with values of $f', \phi', \psi', f, \phi$ and ψ found in previous steps, values of f', ϕ' and ψ' at the end of the interval were predicted by assuming that the previous two values form a geometrical progression with the predicted one; thus the $(s+1)$ th term, f'_{s+1} in a series of values of f' was taken as $f'_{s+1} = (f'_s)^2/f'_{s-1}$. With this assumed value the mean value of f' in the s th interval was taken as $\frac{1}{2}(f'_{s+1} + f'_s)$ and the increment in f was taken as $(0.02)(\frac{1}{2})(f'_{s+1} + f'_s)$. The values of f'_{s+1}, ϕ'_{s+1} and ψ'_{s+1} were then calculated from formulae (14), (7*a*) and (9*a*). If they differed appreciably from the predicted values a second approximation was worked out, replacing the estimated values of f'_{s+1} by this new calculated value. In the early stages of the calculation near $\eta = 1$ two or three approximations were made, but in the later stages the estimated value was so close to the calculated one that the value of f' calculated in this first approximation was used directly in the next stage.

The results are given in table 1 and are shown in the curves of figure 1. These curves and also table 1 show three striking features: (a) the ϕ curve rapidly settles down to a curve which is very nearly a straight line through the origin, (b) the density curve ψ rapidly approaches the axis $\psi = 0$, in fact at $\eta = 0.5$ the density is only 0.007 of the density of the undisturbed atmosphere, (c) the pressure becomes practically constant and equal to $0.436/1.167 = 0.37$ of the maximum pressure. These facts suggest that the solution tends to a limiting form as η decreases in which $\phi = c\eta$, $\phi' = c = \text{constant}$, $f = 0.436$, f', ψ and ψ' become small. Substituting for $\frac{1}{\gamma} \frac{f'}{\psi}$ from (7*a*), (14) becomes

$$\frac{f'}{f} (\eta - \phi)^2 = \gamma \phi' (\eta - \phi) + \frac{3}{2} \gamma \phi - 3\eta + (3 + \frac{1}{2} \gamma) \phi - \frac{2\gamma \phi^2}{\eta}. \quad (19)$$

Dividing by $\eta - \phi$ (19) becomes

$$\frac{f'}{f} (\eta - \phi) = \gamma \phi' - 3 + \frac{2\gamma \phi}{\eta}. \quad (20)$$

If the left-hand side which contains f'/f be neglected the approximate solution of (20) for which ϕ vanishes at $\eta = 0$ is

$$\phi = \eta/\gamma. \quad (21)$$

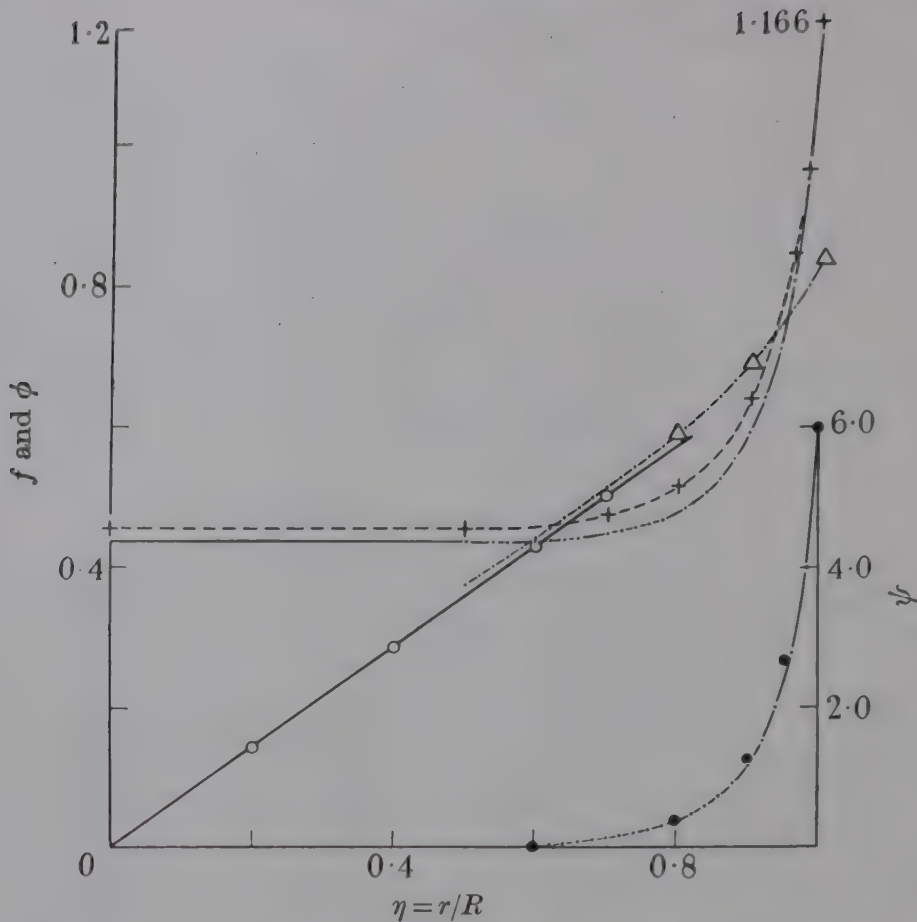


FIGURE 1. - · -, curves f and ψ (step-by-step calculation); -- + --, curve f (approximate formulae). In the other curves the small dots represent the steps of the calculations, the larger symbols represent approximate formulae for: Δ , curve ϕ ; $\circ$, curve $\phi = \eta'\gamma$; $\bullet$, curve ψ .

TABLE 1. STEP-BY-STEP CALCULATION FOR $\gamma = 1.4$

η	f	ϕ	ψ
1.00	1.167	0.833	6.000
0.98	0.949	0.798	4.000
0.96	0.808	0.767	2.808
0.94	0.711	0.737	2.052
0.92	0.643	0.711	1.534
0.90	0.593	0.687	1.177
0.88	0.556	0.665	0.919
0.86	0.528	0.644	0.727
0.84	0.507	0.625	0.578
0.82	0.491	0.607	0.462
0.80	0.478	0.590	0.370
0.78	0.468	0.573	0.297
0.76	0.461	0.557	0.239
0.74	0.455	0.542	0.191
0.72	0.450	0.527	0.152
0.70	0.447	0.513	0.120
0.68	0.444	0.498	0.095
0.66	0.442	0.484	0.074
0.64	0.440	0.470	0.058
0.62	0.439	0.456	0.044
0.60	0.438	0.443	0.034
0.58	0.438	0.428	0.026
0.56	0.437	0.415	0.019
0.54	0.437	0.402	0.014
0.52	0.437	0.389	0.010
0.50	0.436	0.375	0.007

The line $\phi = \eta/\gamma$ is shown in figure 1. It will be seen that the points calculated by the step-by-step method nearly run into this line. The difference appears to be due to the accumulation of errors in calculation.

APPROXIMATE FORMULAE

The fact that the ϕ curve seems to leave the straight line $\phi = \eta/\gamma$ rather rapidly after remaining close to it over the range $\eta = 0$ to $\eta = 0.5$ suggests that an approximate set of formulae might be found assuming

$$\phi = \eta/\gamma + \alpha\eta^n, \quad (22)$$

where n is a positive number which may be expected to be more than, say, 3 or 4. If this formula applies at $\eta = 1$,

$$\frac{1}{\gamma} + \alpha = \frac{2}{\gamma + 1} \quad \text{or} \quad \alpha = \frac{\gamma - 1}{\gamma(\gamma + 1)}; \quad (23)$$

inserting $\phi = \eta/\gamma + \alpha\eta^n$, $\phi' = 1/\gamma + n\alpha\eta^{n-1}$ in (20), the value of f'/f at $\eta = 1$ is $f'/f = \alpha\gamma(n+2)(\gamma+1)/(\gamma-1)$. From (14) and (15b), (16b), (17b) the true value of f'/f at $\eta = 1$ is $\frac{2\gamma^2 + 7\gamma - 3}{\gamma - 1}$. Equating these two forms,

$$n = \frac{7\gamma - 1}{\gamma^2 - 1}. \quad (24)$$

The values of α and n have now been determined to give the correct values of f'/f ; ϕ and ϕ' at $\eta = 1$, ψ' is determined by (9a) so that all the six correct values of f , ϕ , ψ , f' , ϕ' , ψ' are consistent with (22) at $\eta = 1$. Substituting for ϕ from (22) in (20),

$$\frac{f'}{f} = \frac{(n+2)\alpha\gamma^2\eta^{n-2}}{\gamma - 1 - \gamma\alpha\eta^{n-1}}. \quad (25)$$

The integral of (25) which gives the correct value of f at $\eta = 1$ is

$$\log f = \log \frac{2\gamma}{\gamma + 1} - \frac{2\gamma^2 + 7\gamma - 3}{7 - \gamma} \log \left(\frac{\gamma + 1}{\gamma} - \frac{\eta^{n-1}}{\gamma} \right). \quad (26)$$

At $\eta = 0.5$ this gives $f = 0.457$ when $\gamma = 1.4$. The value calculated by the step-by-step integration is 0.436, a difference of 5%.

The approximate form for ψ might be found by inserting the approximate forms for ϕ and ϕ' in (9a). Thus

$$\log \psi = \log \frac{\gamma + 1}{\gamma - 1} - \int_{\eta}^1 \frac{3 + (n+2)\alpha\gamma\eta^{n-1}}{(\gamma - 1)\eta - \alpha\gamma\eta^n} d\eta. \quad (27)$$

Integrating this and substituting for α from (23),

$$\log \psi = \log \frac{\gamma + 1}{\gamma - 1} + \frac{3}{\gamma - 1} \log \eta - 2 \frac{(\gamma + 5)}{7 - \gamma} \log \left(\frac{\gamma + 1 - \eta^{n-1}}{\gamma} \right). \quad (28)$$

When η is small this formula gives

$$\psi = D\eta^{3/(\gamma-1)}, \quad (29)$$

where
$$\log D = \log \frac{\gamma+1}{\gamma-1} - 2 \frac{(\gamma+5)}{7-\gamma} \log \left(\frac{\gamma+1}{\gamma} \right). \quad (30)$$

When $\gamma = 1.4$ (30) gives $D = 1.76$ so that

$$\psi = 1.76\eta^{7.5}. \quad (29a)$$

At $\eta = 0.5$ this gives $\psi = 0.0097$; the step-by-step calculation gives $\psi = 0.0073$. At $\eta = 0.8$ formula (28) gives $\psi = 0.387$, while table 1 gives $\psi = 0.370$. At $\eta = 0.9$ formula (28) gives $\psi = 1.24$, while the step-by-step solution gives 1.18. Some points calculated by the approximate formulae are shown in figure 1.

In the central region of the disturbance the density decreases proportionally to $r^{3/(\gamma-1)}$; the fact that the pressure is nearly constant there means that the temperature increases proportionally to $r^{-3/(\gamma-1)}$. At first sight it might be supposed that these very high temperatures involve a high concentration of energy near the centre. This is not the case, however, for the energy *per unit volume* of a gas is simply $p/(\gamma-1)$ so that the distribution of energy is uniform.

Values of f , ϕ and ψ for $\gamma = \frac{5}{3}$ calculated by the approximate formulae are given in table 2.

TABLE 2. APPROXIMATE CALCULATION $\gamma = 1.666$

η	f	ϕ	ψ
1.00	1.250	0.750	4.00
0.95	0.892	0.680	2.30
0.90	0.694	0.620	1.14
0.80	0.519	0.519	0.63
0.70	0.425	0.445	0.29
0.50	0.379	0.300	0.05
0.00	0.344	0.000	0.00

BLAST WAVE EXPRESSED IN TERMS OF THE ENERGY OF THE EXPLOSION

It has been seen in (18) that $E/\rho_0 A^2$ is a function of γ only. Evaluating the integrals in (18) for the case for $\gamma = 1.4$, and using the step-by-step calculations, it is found that

$$\int_0^1 \eta^2 \phi^2 \psi d\eta = 0.185 \quad \text{and} \quad \int_0^1 \eta^2 f d\eta = 0.187.$$

The kinetic energy of the disturbance is therefore

$$K.E. = 2\pi(0.185)\rho_0 A^2 = 1.164\rho_0 A^2, \quad (31)$$

while the heat energy is

$$H.E. = \frac{4\pi}{(1.4)(0.4)} (0.187)\rho_0 A^2 = 4.196\rho_0 A^2; \quad (32)$$

the total energy is therefore

$$E = 5.36\rho_0 A^2. \quad (33)$$

PRESSURE

The pressure p at any point is

$$p = p_0 R^{-3} f \frac{A^2}{a^2} = R^{-3} f \frac{\rho_0 A^2}{\gamma} = 0.133 R^{-3} E f. \quad (34)$$

The maximum pressure at any distance corresponds with $f = 1.166$ at $r = R$. This is therefore

$$p_{\max.} = 0.155 R^{-3} E. \quad (35)$$

VELOCITY OF AIR AND SHOCK WAVE

The velocity u of the gas at any point is

$$u = R^{-\frac{1}{2}} A \phi = R^{-\frac{1}{2}} E^{\frac{1}{2}} (B \rho_0)^{-\frac{1}{2}} \phi. \quad (36)$$

The velocity of radial expansion of the disturbance is, from (6),

$$\frac{dR}{dt} = A R^{-\frac{1}{2}} = R^{-\frac{1}{2}} E^{\frac{1}{2}} (B \rho_0)^{-\frac{1}{2}}, \quad (37)$$

so that, if t is the time since the beginning of the explosion,

$$t = \frac{2}{5} R^{\frac{5}{2}} (B \rho_0)^{\frac{1}{2}} E^{-\frac{1}{2}} = 0.926 R^{\frac{5}{2}} \rho_0^{\frac{1}{2}} E^{-\frac{1}{2}}, \quad (38)$$

when $\gamma = 1.4$.

The formulae (34) to (38) show some interesting features. Though the pressure wave is conveyed outwards entirely by the air the magnitude of the pressure depends only on ER^{-3} and not on the atmospheric density ρ_0 . The time scale, however, is proportional to $\rho_0^{\frac{1}{2}}$. It is of interest to calculate the pressure-time relationship for a fixed point, i.e. the pressure to which a fixed object would be subjected as the blast wave passed over it. If t_0 is the time since initiation taken for the wave to reach radius R_0 the pressure at time t at radius R_0 is given by

$$\frac{p}{p_1} = \left(\frac{R_0}{R} \right)^3 \frac{f}{[f]_{\eta=1}}, \quad (39)$$

where p_1 is the pressure in the shock wave as it passed over radius R_0 at the time t_0 , R is the radius of the shock wave at time t and $\eta = R_0/R$. $[f]_{\eta=1}$ is the maximum value of f , namely, 1.166 when $\gamma = 1.4$. η is related to t/t_0 through (38) so that

$$\eta = (t_0/t)^{\frac{2}{5}} \quad \text{and} \quad \frac{p}{p_1} = \frac{1}{[f]_{\eta=1}} \left(\frac{t_0}{t} \right)^{\frac{6}{5}} [f]_{\eta=(t_0/t)^{\frac{2}{5}}}. \quad (40)$$

Values of p/p_1 calculated by (40) for $\gamma = 1.4$ are shown in figure 2.

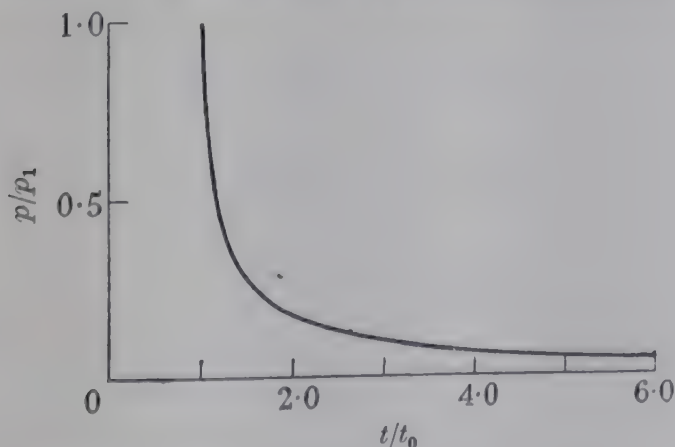


FIGURE 2. Pressure-time curve at a fixed point.

TEMPERATURE

The temperature T at any point is related to the pressure and density by the relationship

$$\frac{T}{T_0} = \frac{p\rho_0}{p_0\rho} = \frac{1.33ER^{-3}}{p_0\psi}, \quad \text{when } \gamma = 1.4. \quad (41)$$

Since f tends to a uniform value 0.436 in the central region ($r < \frac{1}{2}R$) and ψ tends to the value $\psi = 1.76\eta^{7.5}$, T tends to the value

$$\frac{T}{T_0} = \frac{ER^{-3}(0.133)(0.436)}{p_0} \eta^{-7.5} = 0.033 \frac{T_0}{p_0} ER^{-3} \eta^{-7.5}. \quad (42)$$

Thus the temperature near the centre is very high; for instance, when the wave has expanded to such a distance that the pressure in the central region is reduced to atmospheric pressure, $p_0 = (0.133)(0.436)ER^{-3}$, then (42) gives $T/T_0 = \eta^{-7.5}/1.76$ and at $\eta = 0.5$, $\eta^{-7.5} = 181$ so that $T/T_0 = 103$. If $T_0 = 273^\circ$, $T = 27,000^\circ$. The temperature left behind by the blast wave is therefore very high, but the energy density is not high because the density of the gas is correspondingly low.

HEAT ENERGY LEFT IN THE AIR AFTER IT HAS RETURNED TO ATMOSPHERIC PRESSURE

The energy available for doing mechanical work is less than the total heat energy of the air. The heated air left behind by the shock wave can in fact only do mechanical work by expanding down to atmospheric pressure, whereas to convert the whole of the heat energy into mechanical work by adiabatic expansion the air would have to be expanded to an infinite extent till the pressure was zero. After the blast wave has been propagated away and the air has returned to atmospheric pressure it is left at a temperature T_1 , which is greater than T_0 , the atmospheric temperature. The energy required to raise the temperature of air from T_0 to T_1 is therefore left in the atmosphere in a form in which it is not available for doing mechanical work directly on the surrounding atmosphere. This energy, the total amount of which will be denoted by E_1 , is wasted as a blast-wave producer.

The energy so wasted at any stage of the disturbance can be calculated by finding the temperature T_1 to which each element of the blast wave would be reduced if it were expanded adiabatically to atmospheric pressure. If T is the temperature of an element of the blast wave

$$\frac{T}{T_1} = \left(\frac{p}{p_0}\right)^{(\gamma-1)/\gamma} = \left(\frac{A^2}{a^2} f R^{-3}\right)^{(\gamma-1)/\gamma}.$$

Also
$$\frac{T}{T_0} = \frac{p\rho_0}{p_0\rho} = \frac{f}{\psi} \frac{A^2}{a^2} R^{-3},$$

hence
$$\frac{T_1}{T_0} = \frac{f^{1/\gamma}}{\psi} \left(\frac{A^2 R^{-3}}{a^2}\right)^{1/\gamma}. \quad (43)$$

The total heat energy per unit mass of air at temperature T_1 is

$$\frac{T_1}{\gamma-1} \times (\text{gas constant}) = \frac{T_1 p_0}{(\gamma-1) \rho_0 T_0}.$$

The increase in heat energy per unit mass over that which the air contained before the passage of the disturbance is therefore $\frac{p_0}{(\gamma-1)(\rho_0)}\left(\frac{T_1}{T_0}-1\right)$. The increase per unit volume of gas within the disturbed sphere is therefore $\frac{p_0\psi}{\gamma-1}\left(\frac{T_1}{T_0}-1\right)$. Hence from (43) the total energy wasted when the sphere has expanded to radius R is

$$E_1 = 4\pi R^3 \frac{p_0}{\gamma-1} \int_0^1 \left\{ f^{1/\gamma} \left(\frac{A^2 R^{-3}}{a^2} \right)^{1/\gamma} - \psi \right\} \eta^2 d\eta. \tag{44}$$

This expression may conveniently be reduced to non-dimensional form by dividing by the total energy E of the explosion which is related to A by the formula (18). After inserting a^2/γ for p_0/ρ_0 , this gives

$$\frac{E_1}{E} = \frac{4\pi}{B(\gamma-1)\gamma} \left(\frac{a^2}{A^2 R^{-3}} \right) \left[\left(\frac{A^2 R^{-3}}{a^2} \right)^{1/\gamma} \int_0^1 f^{1/\gamma} \eta^2 d\eta - \int_0^1 \psi \eta^2 d\eta \right]. \tag{45}$$

$\frac{4}{3}\pi\rho_0 R^3$ is the total mass of air in the sphere of radius R . This is also $4\pi R^3\rho_0\int_0^1\psi\eta^2d\eta$, so that

$$\int_0^1 \psi \eta^2 d\eta = \frac{1}{3}. \tag{46}$$

The quantity A^2R^{-3}/a^2 is related to the maximum pressure p_1 at the shock wave by the equation

$$y_1 = \frac{p_1}{p_0} = \frac{A^2 R^{-3}}{a^2} [f]_{n=1},$$

where y_1 is the pressure in the shock wave expressed in atmospheres. (46) therefore reduces to

$$\frac{E_1}{E} = \frac{4\pi}{B\gamma(\gamma-1)y_1} [f]_{\eta=1} \left[y_1^{1/\gamma} \int_0^1 \left[\frac{f}{f_{\eta=1}} \right]^{1/\gamma} \eta^2 d\eta - \frac{1}{3} \right]. \tag{47}$$

For $\gamma = 1.4$ numerical integration gives

$$B = 5.36 \text{ (see (18))}, \quad f_{n=1} = 1.166, \quad \int_0^1 f^{1/\gamma} \eta^2 d\eta = 0.219.$$

(47) reduces therefore to

$$\frac{E_1}{E} = \frac{1}{y_1} [0.958 y_1^{1/1.4} - 1.63]. \tag{48}$$

Some values of E_1/E are given in the second line of table 3.

It is clear that E_1/E must continually increase as R increases, and y_1 decreases because the contribution to E_1 due to the air enclosed in the shock-wave surface when its radius is R_2 , say, remains unchanged when this air subsequently expands. A further positive contribution to E_1 is made by each subsequent layer of air included

TABLE 3

y_1 (atm. at shock wave)	10,000	1,000	100	50	20	10*	5*
E_1/E (proportion of energy wasted)	0.069	0.132	0.240	0.281	0.325	0.333*	0.28*
$(E_1 + E_2)/E$	0.096	0.189	0.337	0.393	0.455	—	—

* Formulae inaccurate when $y_1 < 10$.

within the disturbance. The fact that formula (48) gives a value of E_1/E which increases till y_1 is reduced to 10 and then subsequently decreases is due to the inaccuracy of the approximate boundary conditions (15*a*), (16*a*) and (17*a*), which are used to replace the true boundary conditions (15), (16) and (17).

When $\gamma = 1.4$ and $y_1 = 10$ the true value of ρ_1/ρ_0 is 3.8 instead of 6.0 as is assumed, the true value of U^2/a^2 is 8.7 instead of 8.6 and the true value of u_1/U is 0.74 instead of 0.83.

When $y_1 = 5$ the errors are much larger, namely, ρ_1/ρ_0 is 2.8 instead of 6.0, U^2/a^2 is 4.4 instead of 4.3, and u_1/U is 0.64 instead of 0.83. The proportion of the energy wasted, namely, E_1/E , is shown as a function of y_1 in figure 3, y_1 being plotted on a logarithmic scale.

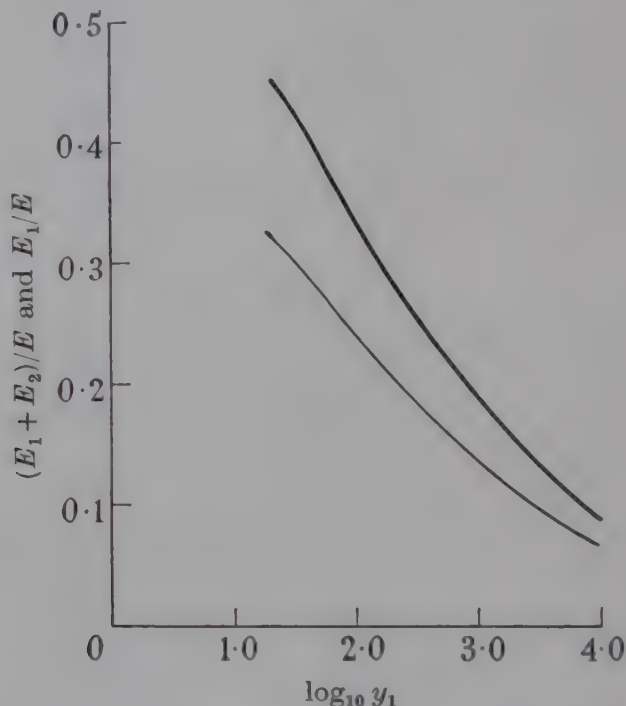


FIGURE 3. Heavy line, $(E_1 + E_2)/E$; thin line E_1/E . $(E_1 + E_2)/E$ is the proportion of the initial energy which is no longer available for doing work in propagation. E_2/E is the work done by heated air expanding against atmospheric pressure (see note added October 1949 (p. 172)).

It will be seen that the limiting value of E_1/E is certainly greater than 0.32, its value for $y_1 = 20$. It is not possible to find out how much greater without tracing the development of the blast wave using laborious step-by-step methods for values of η , less than, say, 10 or 20.

COMPARISON WITH HIGH EXPLOSIVES

The range within which any comparison between the foregoing theory and the blast waves close to actual high explosives can be made is severely limited. In the first place the condition that the initial disturbance is so concentrated that the mass of the material in which the energy is originally concentrated is small compared with the mass of the air involved in the disturbance at any time limits the comparable condition during a real explosion to one in which the whole mass of air involved is several times that of the explosive. In the second place the modified form of the

shock-wave condition used in the analysis is only nearly correct when the rise in pressure at the shock-wave front is several—say at least 5 or 10—atmospheres. In a real explosive this limits the range of radii of shock wave over which comparison could be made to narrow limits. Thus with 10 lb. of C.E.* the radius R at which the weight of explosive is equal to that of the air in the blast wave is 3 ft., while at 3.8 ft. the air is only double the weight of the explosive. The pressure in the blast wave at a radius of 6 ft. was found to be 9 atm., while at 8 ft. it was about 5 atm. It seems, therefore, that in this case the range in which approximate agreement with the present theory could be expected only extends from 3.8 to 6 ft. from the 10 lb. charge.

Taking the energy released on exploding C.E. to be 0.95 kcal./g. the energy released when 10 lb. is exploded is 1.8×10^{14} ergs. If this energy had been released instantaneously at a point as in the foregoing calculations the maximum pressure at distance R given by (35) is

$$p_{\max.} R^3 = (0.155) (1.80 \times 10^{14}) = 2.79 \times 10^{13} \text{ ergs.} \quad (49)$$

Expressed in terms of atmospheres $p_{\max.}$ is identical with y_1 . If R is expressed in feet, (49) becomes

$$y_1 R^3 = \frac{2.79 \times 10^{13}}{(30.45)^3 \times 10^6} = 9.9 \times 10^2. \quad (50)$$

The line representing this relationship on a logarithmic scale is shown in figure 4.

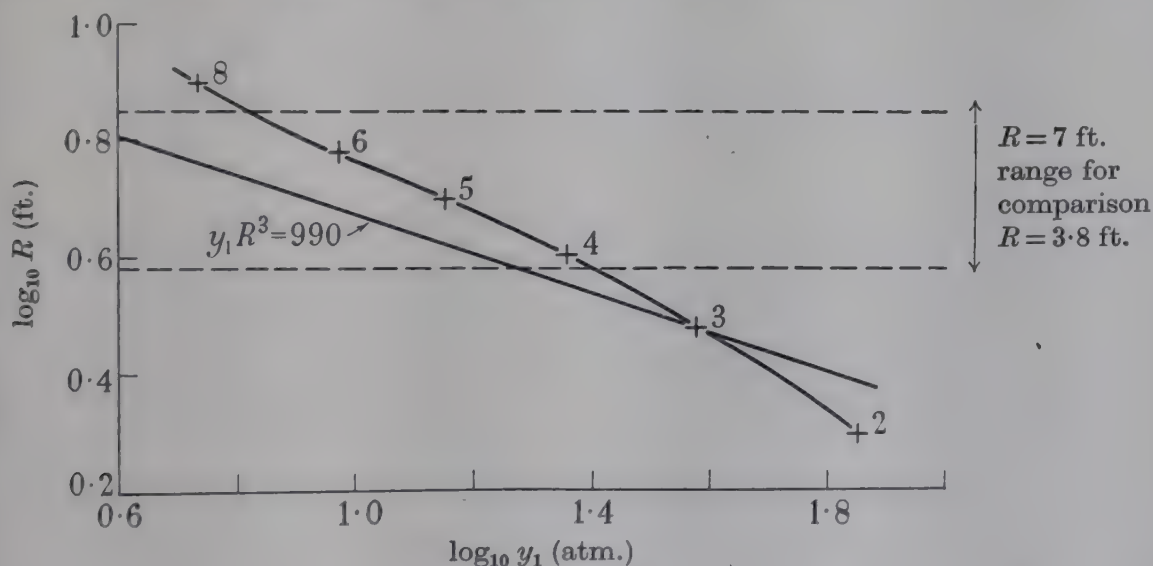


FIGURE 4. Blast pressures near 10 lb. charge of C.E. compared with calculated blast pressures due to instantaneous release of energy of 10 lb. C.E. at a point. The numbers against the points on the curve give distances in feet.

Though no suitable pressure measurements have been made, the maximum pressure in the blast from 10 lb. of C.E. has been found indirectly by observing the velocity of expansion of the luminous zone and, at greater radii, the blast-wave front. These values taken from a curve given in a report on some experiments made by the Road Research Laboratory are given in table 4. The observed values of U in ft./sec. given in column 2 of this table and the values of y_1 (in atmospheres) found from the

* C.E. is the name by which a certain high explosive used in many experiments by the Ministry of Home Security was known.

shock-wave formulae are given in column 3, where they are described as observed values though they were not observed directly. The ‘observed’ values are shown in figure 4. The values of y_1 calculated from (50) are given in column 4.

TABLE 4

R (ft.)	U (ft./sec.)	y_1 (atm.) observed with C.E.	y_1 calculated for concentrated explosion by (50)
8	2350	6.2	—
6	3100	9.3	4.6
5	3800	14.0	7.9
4	4820	22.6	15.5
3	6200	37.5	—
2	8540	71.8	—

Though the observed values are higher than those calculated, it will be noticed that in the range of radii 3.8 to 6 ft., in which comparison can be made, the observed curve is nearly parallel to the theoretical line $y_1 R^3 = 990$. In this range, therefore, the intensity of the shock wave varies nearly as the inverse cube of the distance from the explosion. The fact that the observed values are about twice as great as those calculated on the assumption that the energy is emitted instantaneously at a point may perhaps be due to the fact that the measurements used in table 4 correspond with conditions on the central plane perpendicular to the axis of symmetry of the cylindrical charge used. The velocity of propagation of the luminous zone is greater on this plane and on the axis of symmetry than in other radial directions so that the pressures deduced in column 3 of table 4 are greater than the mean pressures at the corresponding radii.

On the other hand, it has been seen that by the time the maximum pressure has fallen to 20 atm., 32% of the energy has been left behind in the neighbourhood of the concentrated explosive source, raising the air temperature there to very high values. The burnt gases of a real high explosive are at a very much lower temperature even while they are at the high pressure of the detonation wave. Their temperature is still lower when they have expanded adiabatically to atmospheric pressure, so that little heat energy is left in them. To this extent, therefore, a real high explosive may be expected to be more efficient as a blast producer than the theoretical infinitely concentrated source here considered.

Note added, October 1949. The data on which the comparison was based between the pressures deduced by theory and those observed near detonating explosives were obtained in 1940. More recent data obtained at the Road Research Laboratory using a mixture of the two explosives R.D.X. and T.N.T. have been given by Dr Marley. These are given in table 5, which shows the values of U observed for various

TABLE 5. PRESSURE $y_1 p_0$ AT DISTANCE R FROM EXPLOSION OF WEIGHT W OF T.N.T.-R.D.X. MIXTURE

$R/W^{\frac{1}{3}}$ (ft./lb. ^{$\frac{1}{3}$})	0.5	1.0	1.5	2.0	2.5	3.0	3.5
U (thousands ft./sec.)	14.3	11.0	8.4	6.6	5.1	4.0	3.3
y_1 (atm.)	198	117	68.3	42.0	25.1	15.4	10.5
$R/E^{\frac{1}{3}} \times 10^4$ (cm./ergs ^{$\frac{1}{3}$})	5.39	10.8	16.2	21.5	27.0	32.4	37.8

values of $R/W^{\frac{1}{3}}$. R the distance from the explosive is expressed in feet and W its weight in pounds. The third line in table 5 shows the result of deducing y_1 from U using $\gamma = 1.4$ in (16) and $a = 1100$ ft./sec. in (16).

For comparison with the concentrated point-source explosion, the value of $RE^{-\frac{1}{3}}$ expressed in cm. (erg.) $^{-\frac{1}{3}}$ is found by multiplying the figures in line 1, table 5, by $\frac{30.4}{(454)^{\frac{1}{3}} \cdot \frac{1}{(1200 \times 4.2 \times 10^7)^{\frac{1}{3}}}} = 1.078 \times 10^{-3}$. The first factor converts ft. (lb.) $^{-\frac{1}{3}}$ to cm. (g.) $^{-\frac{1}{3}}$, and the second replaces 1 g. by the equivalent energy released by this explosive, namely, 1200 cal. The values of $RE^{-\frac{1}{3}}$ are given in line 4, table 5. In figure 5 values of $\log_{10} y_1$ are plotted against $\log_{10} RE^{-\frac{1}{3}}$, and the theoretical values for a point source of the same energy as the chemical explosive are plotted in the same diagram. Comparing figures 4 and 5 it seems that the more recent shock-wave velocity results are qualitatively similar to the older ones in their relation to the point-source theory. The range of values of y_1 for which comparison between theory and observation might be significant, is marked in figure 5.

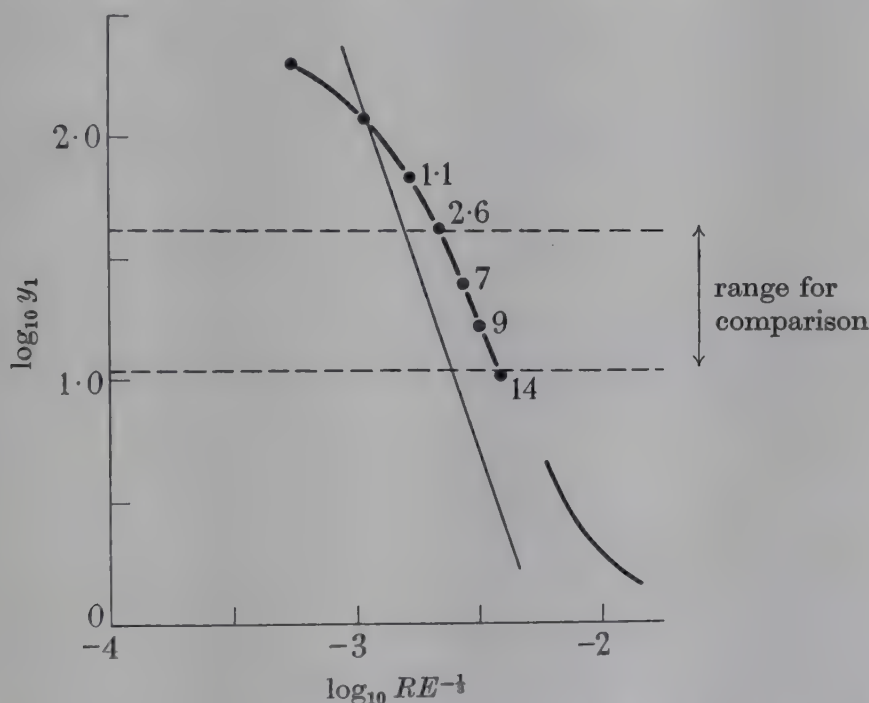


FIGURE 5. Blast pressures near a chemical explosive (R.D.X. + T.N.T.) compared with theoretical pressure for concentrated explosion with same release of energy. Heavy line (upper part) is taken from shock-wave velocity measurements. Heavy line (lower part) is from piezo-electric crystals. Thin line, $y_1 = 0.155E/(p_0R^3)$. The figures against the points represent the ratio of the mass of the air within the shock wave to the mass of the explosive.

It will be seen that the chemical explosive is a more efficient blast producer than a point source of the same energy. The ratio of the pressures in the range of comparison is about 3 to 1. This is more than might be expected in view of the calculation of E_1/E as a function of y_1 which is given in table 3. E_1 is the heat energy which is unavailable for doing mechanical work after expanding to pressure p_0 . Of the remaining energy, $E - E_1$ a part E_2 is used in doing work against atmospheric pressure during the expansion of the heated air. The remaining energy, namely, $E - E_1 - E_2$, is available for propagating the blast wave.

To find E_2 , the work done by unit volume of the gas at radius ηR in expanding to atmospheric pressure is $\left(\frac{T_1 p}{T p_0} - 1\right) p_0$. From (43)

$$\frac{T_1}{T} = \left(\frac{A^2}{a^2} R^{-3} f\right)^{(1-\gamma)/\gamma} \quad \text{and} \quad \frac{p}{p_0} = f \frac{A^2 R^{-3}}{a^2},$$

hence
$$E_2 = 4\pi R^3 p_0 \int_0^1 \left\{ \left(R^{-3} f \frac{A^2}{a^2} \right)^{1/\gamma} - 1 \right\} \eta^2 d\eta, \quad (51)$$

but $\frac{A^2}{a^2 R^3} = \frac{y_1}{(f)_{\eta=1}}$, so that

$$\frac{E_2}{E} = \frac{4\pi R^3 p_0}{E} \left[\left(\frac{y_1}{f_{\eta=1}} \right)^{1/\gamma} \int_0^1 f^{1/\gamma} \eta^2 d\eta - \int_0^1 \eta^2 d\eta \right]. \quad (52)$$

The first integral has already been calculated and found to be 0.219 when $\gamma = 1.4$ (see (47) and (48)). Substituting for $p_{\max}$ from (35),

$$\frac{E_2}{E} = 4\pi(0.155) \left[\frac{0.219 y_1^{(1-\gamma)/\gamma}}{(1.166)^{1/\gamma}} - \frac{1}{3y_1} \right]. \quad (53)$$

Values of $(E_2 + E_1)/E$ have been added as a third line in table 3 and a corresponding curve to figure 3.

REFERENCES

- Taylor, Sir Geoffrey 1946 *Proc. Roy. Soc. A*, **186**, 273.
 Taylor, Sir Geoffrey 1950 *Proc. Roy. Soc. A*, **201**, 175.

The formation of a blast wave by a very intense explosion.

II. The atomic explosion of 1945

BY SIR GEOFFREY TAYLOR, F.R.S.

(Received 10 November 1949)

[Plates 7 to 9]

Photographs by J. E. Mack of the first atomic explosion in New Mexico were measured, and the radius, R , of the luminous globe or 'ball of fire' which spread out from the centre was determined for a large range of values of t , the time measured from the start of the explosion. The relationship predicted in part I, namely, that $R^{\frac{2}{3}}$ would be proportional to t , is surprisingly accurately verified over a range from $R=20$ to 185 m. The value of $R^{\frac{2}{3}}t^{-1}$ so found was used in conjunction with the formulae of part I to estimate the energy E which was generated in the explosion. The amount of this estimate depends on what value is assumed for γ , the ratio of the specific heats of air.

Two estimates are given in terms of the number of tons of the chemical explosive T.N.T. which would release the same energy. The first is probably the more accurate and is 16,800 tons. The second, which is 23,700 tons, probably overestimates the energy, but is included to show the amount of error which might be expected if the effect of radiation were neglected and that of high temperature on the specific heat of air were taken into account. Reasons are given for believing that these two effects neutralize one another.

After the explosion a hemispherical volume of very hot gas is left behind and Mack's photographs were used to measure the velocity of rise of the glowing centre of the heated volume. This velocity was found to be 35 m./sec.

Until the hot air suffers turbulent mixing with the surrounding cold air it may be expected to rise like a large bubble in water. The radius of the 'equivalent bubble' is calculated and found to be 293 m. The vertical velocity of a bubble of this radius is $\frac{2}{3} \sqrt{(g \cdot 29300)}$ or 35.7 m./sec. The agreement with the measured value, 35 m./sec., is better than the nature of the measurements permits one to expect.

COMPARISON WITH PHOTOGRAPHIC RECORDS OF THE FIRST ATOMIC EXPLOSION

Two years ago some motion picture records by Mack (1947) of the first atomic explosion in New Mexico were declassified. These pictures show not only the shape of the luminous globe which rapidly spread out from the detonation centre, but also gave the time, t , of each exposure after the instant of initiation. On each series of photographs a scale is also marked so that the rate of expansion of the globe, or 'ball of fire', can be found. Two series of declassified photographs are shown in figure 6, plate 7.

These photographs show that the ball of fire assumes at first the form of a rough sphere, but that its surface rapidly becomes smooth. The atomic explosive was fired at a height of 100 ft. above the ground and the bottom of the ball of fire reached the ground in less than 1 msec. The impact on the ground does not appear to have disturbed the conditions in the upper half of the globe which continued to expand as a nearly perfect luminous hemisphere bounded by a sharp edge which must be taken as a shock wave. This stage of the expansion is shown in figure 7, plate 8 which corresponds with $t = 15$ msec. When the radius R of the ball of fire reached about 130 m., the intensity of the light was less at the outer surface than in the interior. At

later times the luminosity spread more slowly and became less sharply defined, but a sharp-edged dark sphere can be seen moving ahead of the luminosity. This must be regarded as showing the position of the shock wave when it ceases to be luminous. This stage is shown in figure 8, plate 9, taken at $t = 127$ msec. It will be seen that the edge of the luminous area is no longer sharp.

The measurements given in column 3 of table 1 were made partly from photographs in Mack (1947), partly from some clearer glossy prints of the same photographs kindly sent to me by Dr N. E. Bradbury, Director of Los Alamos Laboratory and partly from some declassified photographs lent me by the Ministry of Supply. The times given in column 2 of table 1 are taken directly from the photographs.

TABLE 1. RADIUS R OF BLAST WAVE AT TIME t AFTER THE EXPLOSION

authority	t (msec.)	R (m.)	$\log_{10} t$	$\log_{10} R$	$\frac{5}{2} \log_{10} R$
strip of small images MDDC 221	0.10	11.1	4.0	3.045	7.613
	0.24	19.9	4.380	3.298	8.244
	0.38	25.4	4.580	3.405	8.512
	0.52	28.8	4.716	3.458	8.646
	0.66	31.9	4.820	3.504	8.759
	0.80	34.2	4.903	3.535	8.836
	0.94	36.3	4.973	3.560	8.901
strip of declassified photographs lent by Ministry of Supply	1.08	38.9	3.033	3.590	8.976
	1.22	41.0	3.086	3.613	9.032
	1.36	42.8	3.134	3.631	9.079
	1.50	44.4	3.176	3.647	9.119
	1.65	46.0	3.217	3.663	9.157
	1.79	46.9	3.257	3.672	9.179
	1.93	48.7	3.286	3.688	9.220
strip of small images from MDDC 221	3.26	59.0	3.513	3.771	9.427
	3.53	61.1	3.548	3.786	9.466
	3.80	62.9	3.580	3.798	9.496
	4.07	64.3	3.610	3.809	9.521
	4.34	65.6	3.637	3.817	9.543
	4.61	67.3	3.688	3.828	9.570
large single photo- graphs MDDC 221	15.0	106.5	2.176	4.027	10.068
	25.0	130.0	2.398	4.114	10.285
	34.0	145.0	2.531	4.161	10.403
	53.0	175.0	2.724	4.243	10.607
	62.0	185.0	2.792	4.267	10.668

To compare these measurements with the analysis given in part I of this paper, equation (38) was used. It will be seen that if the ball of fire grows in the way contemplated in my theoretical analysis, $R^{\frac{5}{2}}$ will be found to be proportional to t . To find out how far this prediction was verified, the logarithmic plot of $\frac{5}{2} \log R$ against $\log t$ shown in figure 1 was made. The values from which the points were plotted are given in table 1. It will be seen that the points lie close to the 45° line which is drawn in figure 1. This line represents the relation

$$\frac{5}{2} \log_{10} R - \log_{10} t = 11.915.$$

(1)

The ball of fire did therefore expand very closely in accordance with the theoretical prediction made more than four years before the explosion took place. This is surprising, because in those calculations it was assumed that air behaves as though γ , the ratio of the specific heats, is constant at all temperatures, an assumption which is certainly not true.

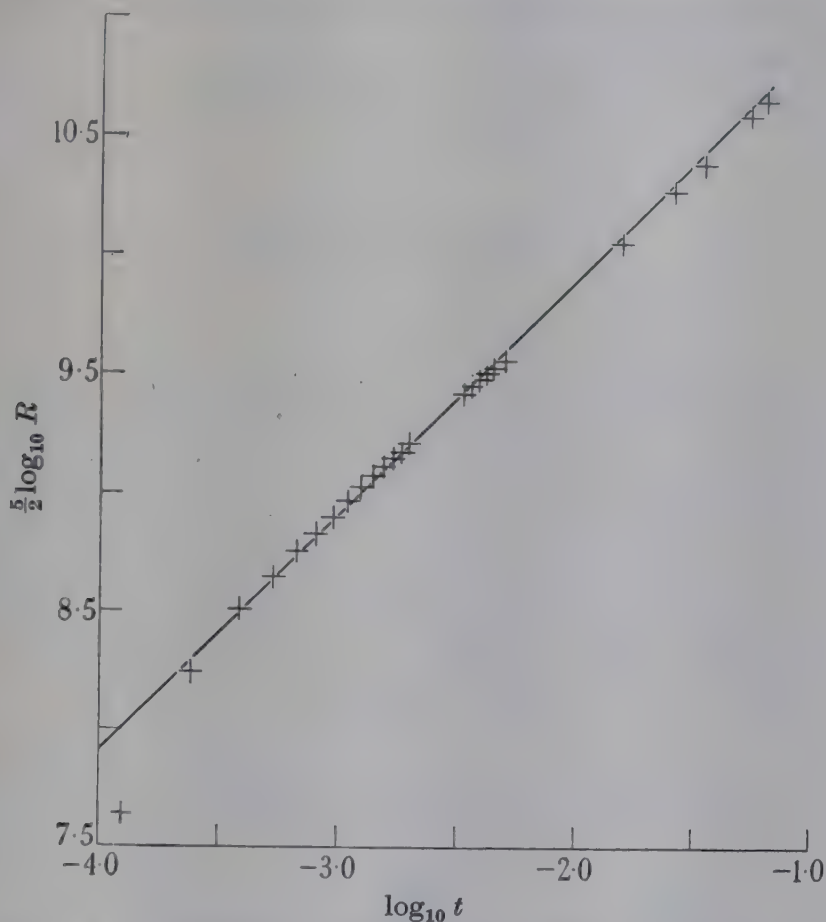


FIGURE 1. Logarithmic plot showing that R^5 is proportional to t .

At room temperatures $\gamma = 1.40$ in air, but at high temperatures γ is reduced owing to the absorption of energy in the form of vibrations which increases C_v . At very high temperatures γ may be increased owing to dissociation. On the other hand, the existence of very intense radiation from the centre and absorption in the outer regions may be expected to raise the apparent value of γ . The fact that the observed value of $R^5 t^{-2}$ is so nearly constant through the whole range of radii covered by the photographs of the ball of fire suggests that these effects may neutralize one another, leaving the whole system to behave as though γ has an effective value identical with that which it has when none of them are important, namely, 1.40.

CALCULATION OF THE ENERGY RELEASED BY THE EXPLOSION

The straight line in figure 1 corresponds with

$$R^5 t^{-2} = 6.67 \times 10^2 (\text{cm.})^5 (\text{sec.})^{-2}. \quad (2)$$

The energy, E , is then from equation (18) of part I

$$E = \rho_0 A^2 \left\{ 2\pi I_1 + \frac{4\pi}{\gamma(\gamma-1)} I_2 \right\}, \quad (3)$$

where

$$I_1 = \int_0^1 \psi \phi^2 \eta^2 d\eta$$

(4)

and

$$I_2 = \int_0^1 f \eta^2 d\eta,$$

(5)

where f , ϕ and ψ are non-dimensional quantities proportional to pressure, velocity and density which are defined in part I.

A is found by integrating equation (6) of part I, so that

$$A = \frac{2}{5} R^{\frac{1}{2}} t^{-1}.$$

(6)

Writing

$$K = \frac{4}{25} \left\{ 2\pi I_1 + \frac{4\pi}{\gamma(\gamma-1)} I_2 \right\},$$

(7)

$$E = K \rho_0 R^5 t^{-2}.$$

(8)

It was shown in part I that I_1 and I_2 are functions of γ only. For $\gamma = 1.40$, their values were found to be $I_1 = 0.185$, $I_2 = 0.187$, using step-by-step methods for integrating the equations connecting f , ϕ and ψ . Using the approximate formulae (22) to (30) of part I, values of f , ϕ and ψ were calculated in part I for $\gamma = 1.667$ and are there given in table 2 of part I. Further calculations have now been made for $\gamma = 1.20$ and $\gamma = 1.30$ using the approximate formulae. The results for $\gamma = 1.30$ are given in table 2 and are shown in figure 2. The corresponding values of I_1 , I_2 and K are given in table 3; K is shown as a function of γ in figure 3.

TABLE 2. CALCULATION FOR $\gamma = 1.30$ USING FORMULAE (22) to (30) OF PART I

η	f	ϕ	ψ	T/T_1
1.0	1.130	0.869	7.667	1.00
0.98	0.896	0.833	4.534	1.34
0.96	0.772	0.801	2.989	1.75
0.94	0.666	0.772	2.067	2.19
0.92	0.606	0.745	1.454	2.73
0.90	0.563	0.721	1.058	3.61
0.85	0.499	0.669	0.509	6.76
0.80	0.468	0.623	0.255	12.5
0.75	0.453	0.580	0.128	24.0
0.70	0.445	0.540	0.063	48.0
0.65	0.441	0.501	0.029	103
0.60	0.439	0.462	0.013	229
0.55	0.438	0.423	0.006	—
0.50	0.438	0.386	0.002	—
0.45	0.438	0.347	0.001	—
0.40	0.438	0.309	0.000	—

ENERGY OF THE FIRST ATOMIC EXPLOSION IN NEW MEXICO

Having determined $R^5 t^{-2}$ and assuming that ρ_0 may be taken as 1.25×10^{-3} g./cm.³, the figures in table 3 were used to determine E from equations (8) and (2). Different values are found for different assumed values of γ . These are given, expressed in ergs, in line 5 of table 3. It has become customary to describe large explosions by stating the weight of T.N.T. which would liberate the same amount of energy. Taking 1 g.

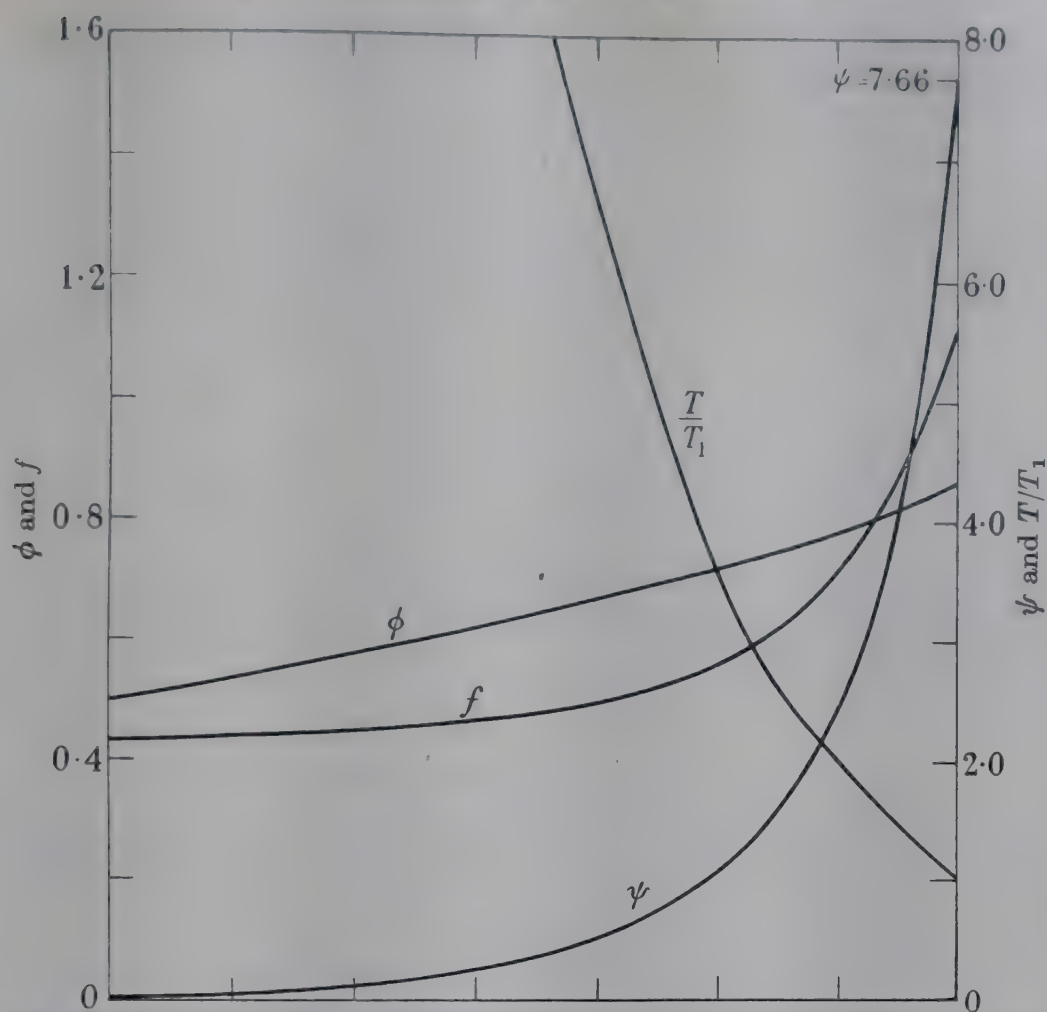


FIGURE 2. Distribution of radial velocity ϕ , pressure f , density ψ and temperature T/T_1 for $\gamma = 1.30$ expressed in non-dimensional form.

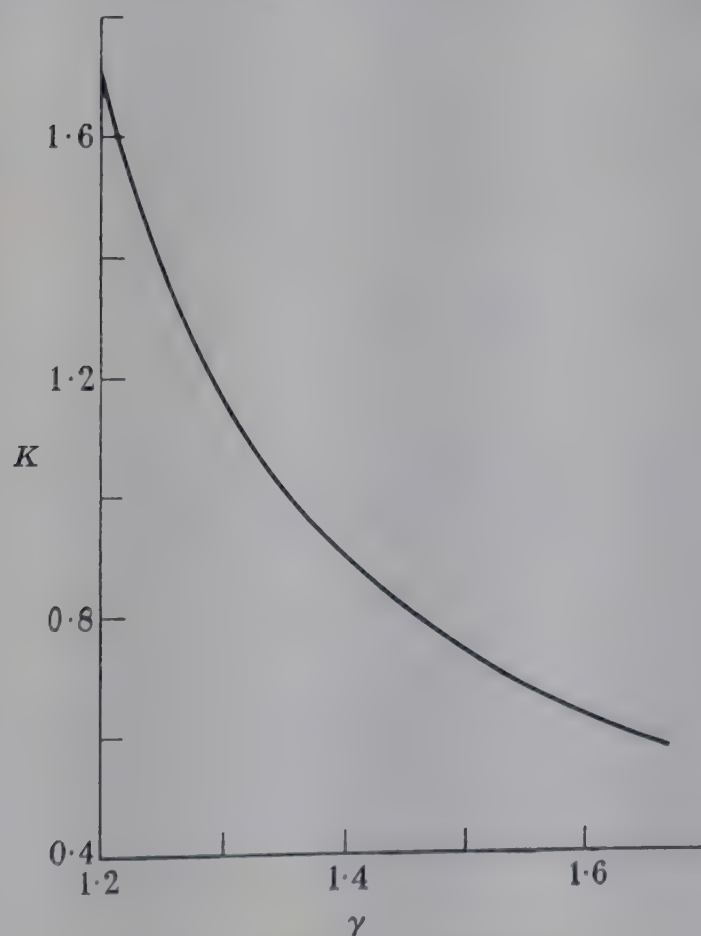


FIGURE 3. Variation of K with γ .

of T.N.T. as liberating 1000 calories or 4.18×10^{10} ergs, 1 ton will liberate 4.25×10^{16} ergs. The T.N.T. equivalents found by dividing the figures given in line 5 of table 3 by 4.25×10^{16} are given in line 6.

TABLE 3. CALCULATED CONSTANTS USED IN DETERMINING THE ENERGY OF THE EXPLOSION WITH A RANGE OF ASSUMED VALUES FOR γ

γ	1.20	1.30	1.40	1.667
I_1	0.259	0.221	0.185	0.123
I_2	0.175	0.183	0.187	0.201
K	1.727	1.167	0.856	0.487
$E \times 10^{-20}$ erg	14.4	9.74	7.14	4.06
T.N.T. equivalent (tons)	34,000	22,900	16,800	9,500

It will be seen that if $\gamma = 1.40$, the T.N.T. equivalent of the energy of the New Mexico explosion, or more strictly that part of the energy which was not radiated outside the ball of fire, was *16,800 tons*.

AN ALTERNATIVE POSSIBILITY

If the effect of radiation, which cannot be estimated, is disregarded, a mean value of γ might be taken which is appropriate to the temperature calculated to correspond with the mean value of R in the range over which $R^{\frac{1}{2}}t^{-1}$ is nearly constant. The least value of R which lies near the line in figure 1 is $R = 20$ m. and the greatest is $R = 185$ m. The mean value is therefore approximately 100 m. It will be found that the value of γ appropriate to the temperature behind the shock wave at 100 m. is about 1.3. The pressure at any point is from equations (1), (6) and (12) of part I

$$p = p_0 R^{-3} f_1 = p_0 R^{-3} \frac{A^2}{a^2} f = \frac{4}{25} \frac{\rho_0}{\gamma} R^{-3} f(R^5 f^{-2}).$$

For $\gamma = 1.3$, the value of f behind the detonation front is 1.13, so that

$$p = 9.3 \times 10^{22} \rho_0 R^{-3}.$$

The temperature T_1 at that point is given by

$$\frac{T_1}{T_0} = \frac{p}{\rho} \frac{\rho_0}{p_0},$$

where T_0 is the undisturbed atmospheric temperature. When

$$\gamma = 1.3, \quad \rho_0/\rho = 1/\psi = 1/7.667, \quad \text{and if} \quad \rho_0 = 0.00125, \quad p/p_0 = 7.43 \times 10^{13} R^{-3}.$$

Thus

$$\frac{T_1}{T_0} = 0.97 \times 10^{13} R^{-3}. \quad (9)$$

At $R = 100$ m., $T_1/T_0 = 9.7$, so that if $T_0 = 15^\circ \text{C} = 288^\circ \text{K}$, $T_1 = 2800^\circ \text{K}$.

Values of C_p at temperatures up to 5000°K have been calculated for nitrogen by Johnston & Davis (1934) and for oxygen by Johnston & Walker (1935). At 2800°K , C_p is given for nitrogen as 8.82 and for oxygen as 9.43, so that for air $C_p = 8.92$. Since there is little dissociation at that temperature, it seems that $C_v = C_p - R = 6.92$

and $\gamma = \frac{8.92}{6.92} = 1.29$. Thus the use of $\gamma = 1.30$ for calculating the temperature at $R = 100$ m. is justified when effects of radiation are neglected.

Using the curve, figure 3, it seems that the value of K appropriate to $\gamma = 1.29$ is 1.21. Using $\rho_0 = 0.00125$ and $R^5 t^{-2} = 6.67 \times 10^{23}$, equation (8) gives $E = 1.01 \times 10^{21}$ ergs and the T.N.T. equivalent is 23,700 tons.

Of these two alternative estimates, it seems that the first, namely, 16,800 tons, is the more likely to be accurate.

SOME DYNAMICAL FEATURES OF THE ATOMIC EXPLOSION

It will be seen in figure 1 of part I that if $\gamma = 1.4$ the air density is a maximum at the shock wave front where it reaches six times atmospheric density. Within the shock wave, the density falls rapidly till at a radius of about $0.6R$ it is nearly zero. Within the radius $0.6R$ the gas has a radial velocity which is proportional to the distance from the centre, a very high temperature, and a uniform pressure about 0.43 time the maximum pressure.

The maximum pressure at the shock front is found by inserting the value of E from line 5 of table 3 in the formula (35) of part I. The pressure—expressed in atmospheres—is

$$\frac{p}{p_0} = 0.155R^{-3}(7.14 \times 10^{20}) \times 10^{-6} = 1.11 \times 10^{14}R^{-3}. \quad (10)$$

At $R = 30$ m. this is 4100 atm., or 27 tons/sq.in. At $R = 100$ m. the pressure would be 1 ton/sq.in. These pressures are much less than would act on rigid bodies exposed to such blasts, but the pressures on obstacles depend on their shape so that no general statement can be made on this subject.

The temperature rises rapidly as the centre is approached, in fact the ratio T/T_1 , T_1 being the temperature just inside the shockwave, is equal to

$$\frac{p_1}{p} \left(\frac{\rho_1}{\rho} \right) = \frac{p}{p_0} \frac{p_0 \rho_1 \rho_0}{p_1 \rho_0 \rho} = \left(\frac{f}{f_{\eta=1}} \right) \left(\frac{\psi_{\eta=1}}{\psi} \right). \quad (11)$$

The values of T/T_1 for $\gamma = 1.3$, namely $\frac{7.66 f}{1.13 \psi}$, are given in the fifth column of table 2, and are shown in figure 2.

The initial rate of rise of air from the seat of the explosion

When the shock wave had passed away from the ball of fire, it left a cloud of very hot air which then rapidly rose. Mack (1947, p. 37) gives a rough picture of the process in a series of diagrams representing the outlines of the boundary of the heated region, so far as his photographs could define them, at successive times from $t = 0.1$ to $t = 15.0$ sec. It is not possible to know exactly what these outlines represent, though in the later numbers of the series they seem to show the limits of the region to which dust thrown off from the ground and sucked into the ascending column of air has penetrated. This dust rapidly expands into a roughly spherical shape owing

to turbulent diffusion or convection currents in the central region. The radius of the outer edge of the glowing region is not the same in all photographs taken nearly simultaneously, but the height of its apparent centre seems to be consistent when photographs taken simultaneously from different places are compared.

The heights, h , of the top of the illuminated column and their radii, b , were determined, so far as was possible, from Mack's published photographs. These rough measurements are given in table 4. The height of the centre of the glowing area is taken as $h - b$, and points corresponding with those given in table 4 are plotted in figure 4. It will be seen that the centre of the glowing volume seems to rise at a regular rate. The line drawn in figure 4 corresponds with a vertical velocity of

$$U = 35 \text{ m./sec.}$$

(12)

TABLE 4. HEIGHT, h , AND RADIUS, b , OF THE GLOWING REGION FROM 3½ TO 15 SEC. AFTER THE EXPLOSION

authority	t (sec.)	b (radius) (m.)	h (height of top)	$h - b$ (height of centre)
Mack MDCC 221, sketches on p. 37	3.5	160	375	215
	8.0	240	688	448
	10.0	300	810	510
	15.0	360	1060	700
photograph on p. 38	14.8	550	1200	650

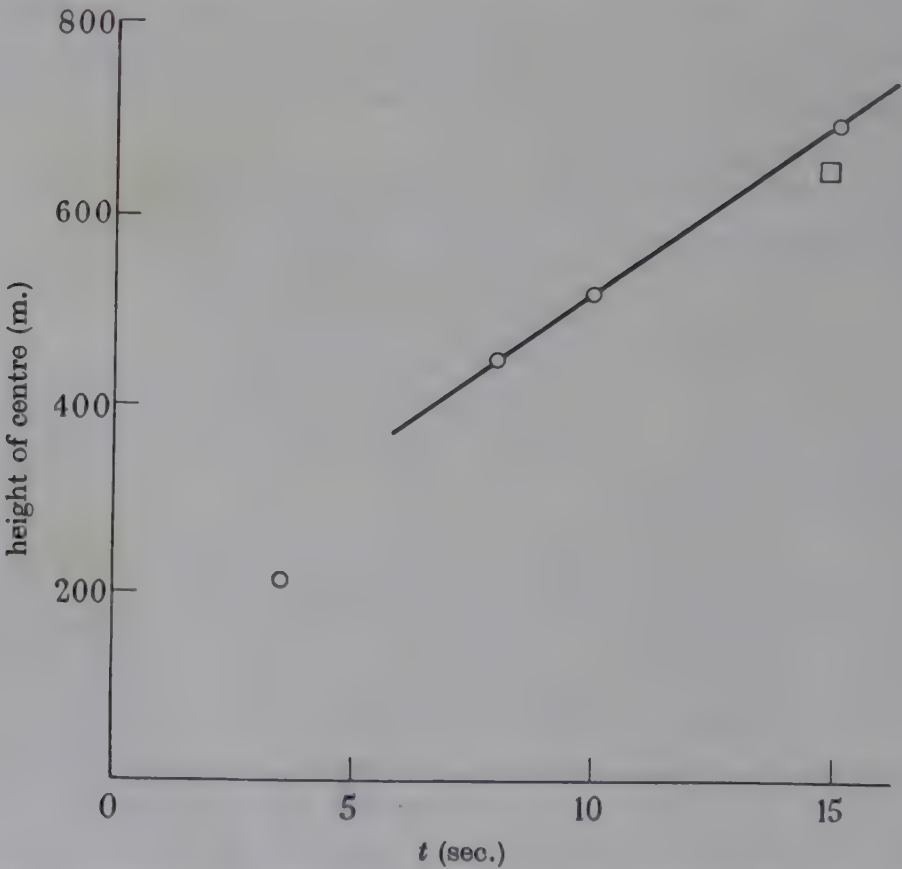


FIGURE 4. Height of centre of glowing region from 3½ to 15 sec. after the explosion. ○ from diagram p. 37, □ from photograph p. 38 of MDCC 221.

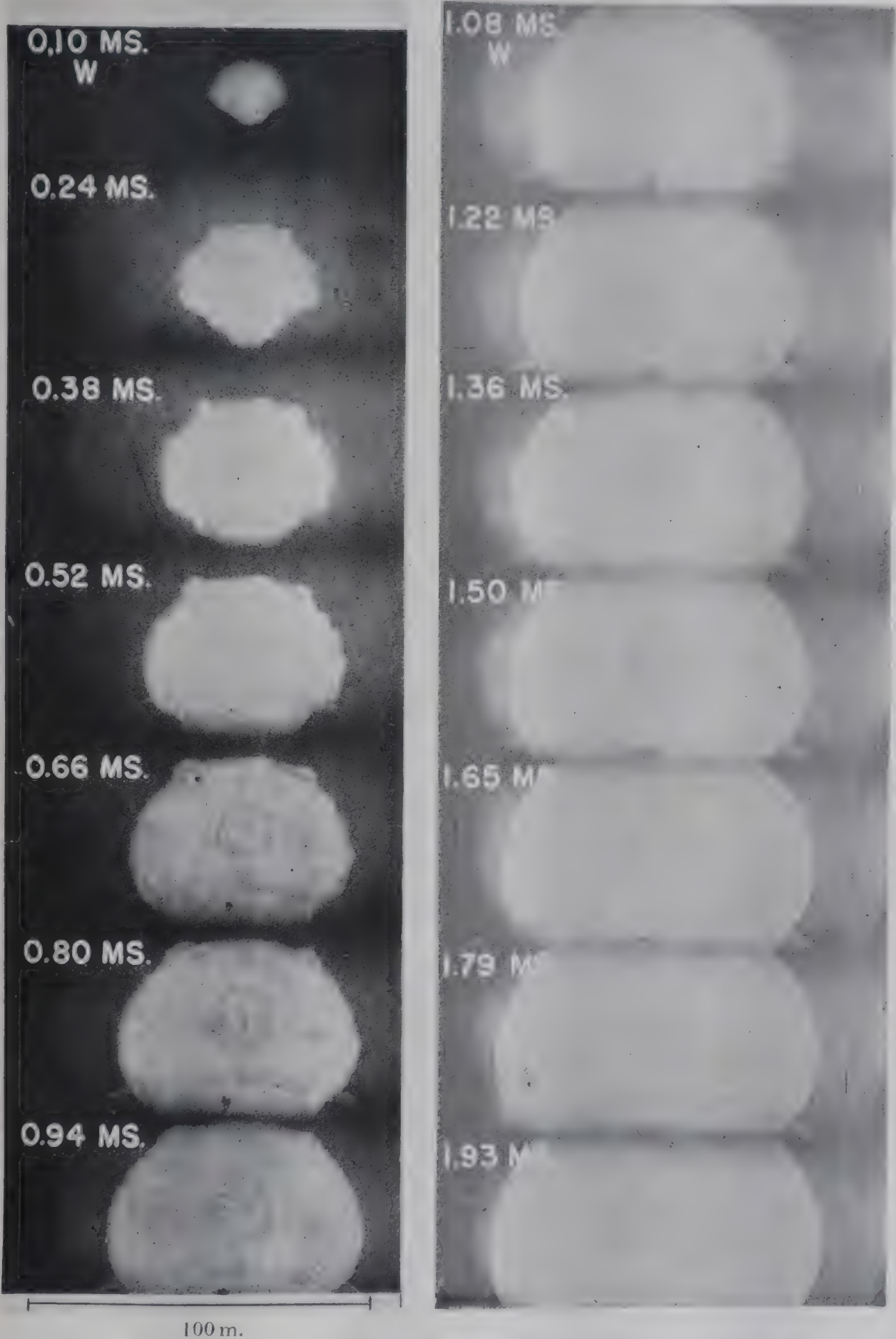


FIGURE 6. Succession of photographs of the 'ball of fire' from $t = 0.10$ msec. to 1.93 msec.



FIGURE 7. The ball of fire at $t = 15$ msec., showing the sharpness of its edge.

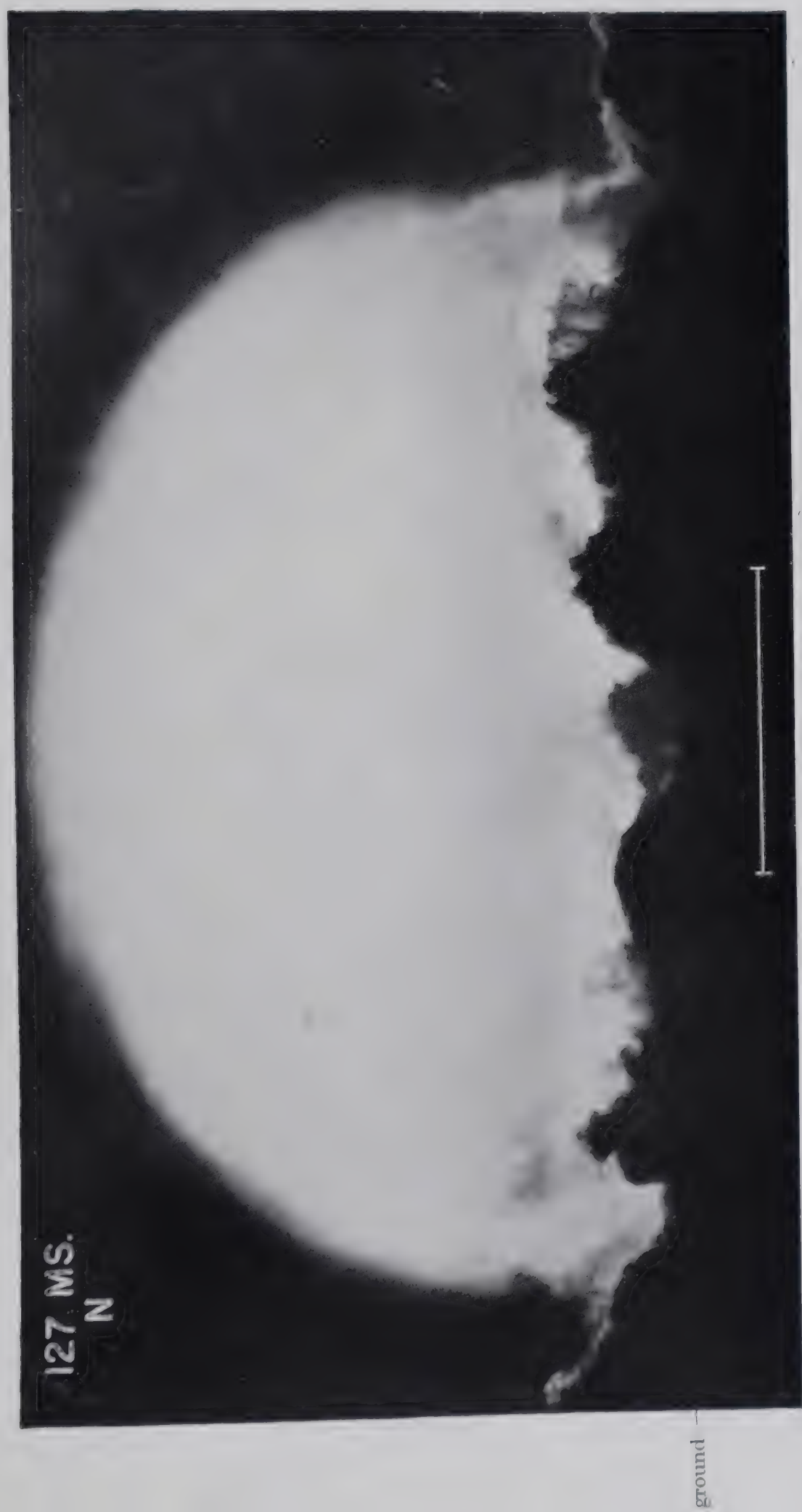


FIGURE 8. The glowing volume $t = 127$ msec. showing an indefinite edge and hemispherical shape.

Distribution of air density after the explosion

To give a dynamical description of the rise of hot air from the seat of the explosion it is first necessary to know the distribution of density immediately after the shock wave has passed away and left hot air at atmospheric pressure. Formulae are given in part I for the temperature ultimately attained by air which passed through the shock wave when its pressure was $y_1 p_0$, but the position at which this air comes to rest when atmospheric pressure was attained was not discussed. If the ground had not obstructed the blast wave the distribution of temperature would evidently be spherical. It has been pointed out, however, that the shape of the upper half of the blast wave has not been affected by the presence of the ground. The same thing seems to be approximately true of the temperature distribution, for Mack publishes a photograph showing the luminous volume at $t = 0.127$ sec. when the shock wave had moved well away from the very hot area. This is reproduced in figure 8, plate 9. It will be seen that the glowing air occupies a nearly hemispherical volume, the bottom half of the sphere being below the ground. It seems that it may be justifiable to assume that most of the energy associated with the part of the blast wave which strikes the ground is absorbed there. In that case we may neglect the effect of shock waves reflected from the ground and consider the temperature distribution as being that calculated in part I for an unobstructed wave. With this assumption the distribution of density will be calculated.

The following symbols will be used:

T_0, p_0, ρ_0 , the temperature, pressure and density in the undisturbed air,

$T, y_1 p_0, \rho, R$, the temperature, pressure, density and radius at the shock wave,

T_1, ρ_1, r , the temperature, density and radius after the pressure has become atmospheric.

From (43), part I,

$$\frac{T}{T_1} = y_1^{(\gamma-1)/\gamma} = \left(\frac{A^2 f R^{-3}}{a^2} \right)^{(\gamma-1)/\gamma} \quad (13)$$

and

$$\frac{T}{T_0} = \frac{f}{\psi} \frac{A^2}{a^2} R^{-3}, \quad (14)$$

so that

$$\frac{T_1}{T_0} = \frac{T}{T_1} \frac{T_0}{T} = \frac{1}{\psi} \left(\frac{A^2 f R^{-3}}{a^2} \right), \quad (15)$$

and since

$$\frac{A^2}{a^2} = \frac{4E}{25K\gamma p_0}, \quad (16)$$

$$\frac{T_1}{T_0} = \beta \left(\frac{R}{r_0} \right)^{-3/\gamma},$$

where

$$\beta = \frac{1}{\psi} \left(\frac{4f}{25K\gamma} \right)^{1/\gamma}, \quad (17)$$

and r_0 is the length defined by $r_0^3 = E/p_0$. (18)

r_0 is introduced in order to make the equations non-dimensional, ψ and f have their values at the shock wave front. When $\gamma = 1.40, f = 1.167, \psi = 6.0; K = 0.856$ (see table 3), so that

$$\beta = 0.044, \quad (19)$$

(16) may be written

$$\left(\frac{R}{r_0}\right)^3 = \beta^\gamma \left(\frac{T_0}{T_1}\right)^\gamma = 0.1265 \left(\frac{T_0}{T_1}\right)^\gamma \quad \text{when } \gamma = 1.40. \tag{20}$$

To find the values of r and T_1/T_0 corresponding with a given value of y_1 the equation of continuity must be used. This is

$$\rho_0 R^2 dR = \rho_1 r^2 dr, \tag{21}$$

and since $T_0/T_1 = \rho_1/\rho_0$, (21) can be integrated after substitution from (20), thus

$$\left(\frac{r}{r_0}\right)^3 = \beta^\gamma \left(\frac{\gamma}{\gamma-1}\right) \left(\frac{T_0}{T_1}\right)^{\gamma-1}, \tag{22}$$

and from (35) of part I and (20)

$$\left(\frac{R}{r_0}\right)^3 = \frac{0.155}{y_1} = \beta^\gamma \left(\frac{T_0}{T_1}\right)^\gamma. \tag{23}$$

Eliminating T_0/T_1 from (22) and (23) and using values appropriate to $\gamma = 1.40$

$$\left(\frac{r}{r_0}\right)^3 = 0.0906 y_1^{-0.2857} \tag{24}$$

and

$$\frac{\rho_0}{\rho_1} = \frac{T_1}{T_0} = 0.167 y_1^{0.7143}. \tag{25}$$

Values of R/r_0 , ρ_1/ρ_0 , r/r_0 calculated for a range of values of y , are given in table 5.

TABLE 5. DENSITY ρ_1 AT RADIUS r EXPRESSED AS A PROPORTION OF ρ_0 , THE UNDISTURBED DENSITY

y_1 (atm.)	r/r_0	ρ_1/ρ_0 uncorrected	C	ρ_1/ρ_0 corrected	R/r_0
10	(0.360)	1.155	0.753	0.873	0.249
20	0.337	0.704	0.858	0.605	0.198
30	0.325	0.527	0.900	0.475	0.173
40	0.316	0.430	0.923	0.397	0.157
60	0.304	0.321	0.947	0.305	0.137
100	0.289	0.223	0.968	0.216	0.116
500	0.247	0.071	—	0.071	0.068
5000	0.200	0.013	—	0.013	0.031

It has been pointed out that T_1/T_0 contains two factors, T/T_0 and T_1/T . T/T_0 represents the temperature change through the shock wave and equal to $\rho_0 y_1/\rho$. In calculating this the approximate expression $\frac{\rho_0}{\rho} = \frac{\gamma-1}{\gamma+1}$ was used instead of the true value

$$\frac{\rho_0}{\rho} = \frac{\gamma+1+(\gamma-1)y_1}{\gamma-1+(\gamma+1)y_1}. \tag{26}$$

The proportional error in calculating T/T_0 for a given values of y_1 is therefore equal to the proportional error in using $(\gamma-1)/(\gamma+1)$ instead of the correct expression for ρ_0/ρ . The second factor, T_1/T , which represents the reduction in temperature for

a given expansion ratio is correct, so that correct values of ρ/ρ_0 for a given value of y_1 can be found by multiplying the figures given in column 3 of table 5 by a correcting factor

$$C = \frac{\gamma - 1 + (\gamma + 1)y_1}{\gamma + 1 + (\gamma - 1)y_1} \cdot \frac{\gamma - 1}{\gamma + 1}. \quad (27)$$

Values of C are given in column 4, table 5, and the corrected values of ρ_1/ρ_0 in column 5. It must be pointed out that though the figures in column 5 are correct, the values obtained for r/r_0 when y_1 is less than about 40 are subject to an appreciable error owing to using approximate values of f , ϕ and ψ at the shock front.

The variation of ρ_1/ρ_0 with r/r_0 is shown in figure 5. It will be seen that the density is very small when $r/r_0 < 0.2$, that it begins to rise steeply at about $r/r_0 = 0.28$, and that it has nearly attained atmospheric density when $r/r_0 = 0.36$.

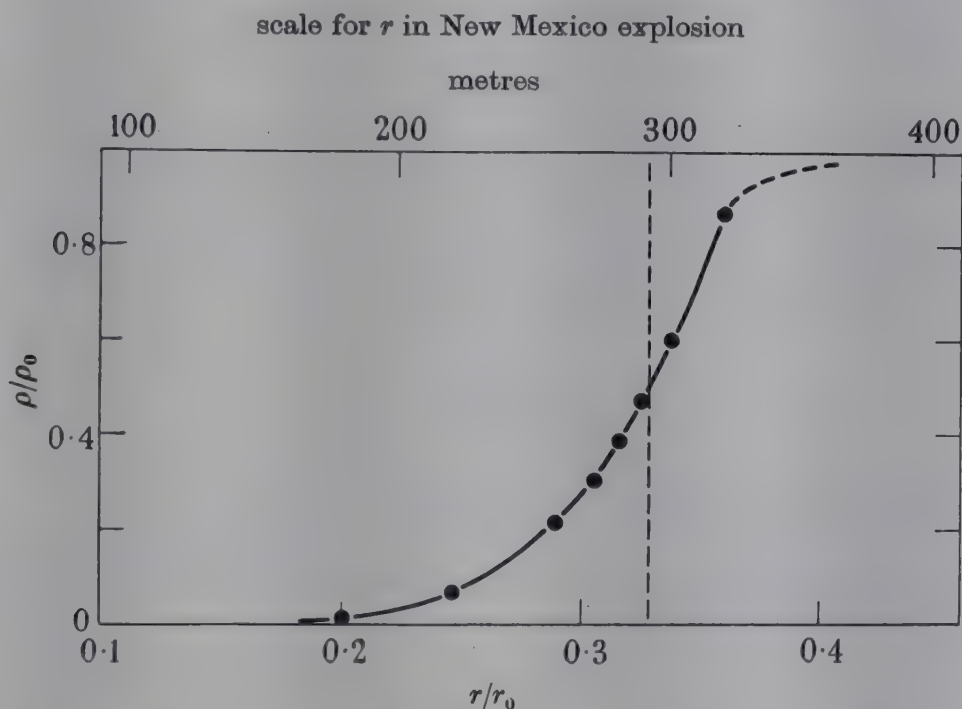


FIGURE 5. Distribution of density after the explosion. The upper scale corresponds with calculations using the value of E given in table 3 for $\gamma = 1.40$.

Calculation of the rate of rise of the heated air

Though it would be difficult to calculate the effect of buoyancy on a fluid with the density distribution shown in figure 5, the rise of bubbles in water, when the change from very light to heavy fluid is discontinuous, has been studied. It has been shown both experimentally and theoretically (Davies & Taylor 1950) that the vertical velocity U of a large bubble is related to a , the radius of curvature of the top of the bubble, by the formula

$$U = \frac{2}{3} \sqrt{ga}. \quad (28)$$

It seems worth while to compare the observed rate of rise of the air heated by the New Mexico explosion with that of a large bubble in water, and for that purpose it is necessary to decide on the radius of a sphere of zero density which might be expected to be comparable with air having the density distribution of figure 5. The

simplest guess is to take the radius as the value of r for which $\rho/\rho_0 = \frac{1}{2}$, and in figure 5 this corresponds with the broken line for which

$$r/r_0 = 0.328. \quad (29)$$

In the New Mexico explosion the best estimate of E was that corresponding with the measured rate of expansion of the ball of fire, assuming $\gamma = 1.40$. This is given in table 3, namely $E = 7.14 \times 10^{20}$ ergs.

Using this value and $p_0 = 10^6$ dynes/sq.cm.

$$r_0 = (7.14 \times 10^{14})^{\frac{1}{3}} = 8.9 \times 10^4 = 894 \text{ m.} \quad (30)$$

The radius chosen for comparison with a bubble rising in water is therefore

$$r = 0.328 \times 894 = 293 \text{ m.} \quad (31)$$

The predicted velocity of rise is therefore

$$U = \frac{2}{3} \sqrt{(981 \times 2.93 \times 10^4)} = 35.7 \text{ m./sec.} \quad (32)$$

Comparing this with the observed value of the vertical velocity of the centre of the glowing volume, namely, 35 m./sec., it will be seen that the agreement is better than the nature of the measurements would justify one in expecting. A far less good agreement would justify a belief that the foregoing dynamical picture of the course of events after the atomic explosion is essentially correct.

REFERENCES

- Davies, R. M. & Taylor, Sir G. 1950 *Proc. Roy. Soc. A*, **200**, 375.
 Johnston, H. L. & Davis, C. O. 1934 *J. Amer. Chem. Soc.* **56**, 271.
 Johnston, H. L. & Walker, M. K. 1935 *J. Amer. Chem. Soc.* **57**, 682.
 Mack, J. E. 1947 *Semi-popular motion picture record of the Trinity explosion*. MDDC 221.
 U.S. Atomic Energy Commission.

On the reproduction and ruling of diffraction gratings

BY SIR THOMAS MERTON, TREAS.R.S.

(Received 7 December 1949)

New methods of copying ruled diffraction gratings are described. Copies of gratings can be made in hardened gelatine, and the technique of preparing such gratings and their advantages over pyroxyline replicas are discussed. Replicas on flat glass coated with gelatine can also be made from gratings ruled on cylinders. Cylinders ruled on a lathe are not sufficiently accurate but a method is described by means of which periodic errors can be eliminated.

In recent years there has been a greatly increased demand for diffraction gratings. The advances in infra-red spectroscopy during the last ten years have created a need for gratings of a type specially suited to the study of the infra-red spectrum, and there is also a requirement of very large gratings with relatively coarse rulings for use in conjunction with astronomical reflectors. In addition to this there has been an expansion in the use of spectrographic analysis for industrial purposes with a very noticeable tendency to substitute grating instruments for the quartz spectrographs which were more commonly used twenty years ago. The ruling of diffraction gratings with 15,000 or 20,000 lines to the inch and of the required accuracy is one of the most difficult of engineering achievements and has been accomplished successfully by very few investigators. It is doubtful whether the gratings ruled by Rowland nearly seventy years ago have ever been surpassed in quality.

It is natural therefore that many attempts should have been made to prepare copies or replicas of original rulings. One of the commonest methods of copying gratings is based on the method of Thorpe (Brit. Patent), which consists of making a cast of the grating in pyroxyline. A number of variations of this process have been developed by different investigators, but in essence it consists in pouring a solution of pyroxyline in amyl acetate on the surface of the grating to be copied and allowing the solvent to evaporate. Another method consists in 'spreading' a viscous solution (15 %) of cellulose acetate in acetone on the grating with a straight edge using wires of appropriate thickness as distance pieces. The grating is now put into a dish of water if pyroxyline is used, and after a time the pellicle becomes detached. If cellulose acetate is used the pellicle can be stripped without wetting. The pellicle can then be mounted on a flat sheet of glass in a variety of ways, either by coating the glass with a layer of gelatine and mounting the pellicle with the rulings uppermost on the wet and swollen gelatine, or by squeegeeing the pellicle with the rulings down in contact with uncoated glass covered with water, the edges of the pellicle in this case being ringed with a solution of gelatine or some other cement to keep the pellicle in position. In the first-mentioned method the water in the swollen gelatine evaporates through the pellicle and holds it in position on drying, and in the latter method the water between the pellicle and the supporting glass evaporates in the same way, the diffraction spectrum appearing in full brilliance only after the water has evaporated. These methods are fully described by Wallace (1905).

These replicas, which can be converted into reflexion gratings by aluminizing, have never proved to be a completely effective substitute for the original ruled gratings. There are two reasons for this. In the first place it is vital that the pellicle should not be distorted in the process of removing it from the original grating and transferring it to its final support. In the second the pellicle is never of uniform thickness. The distortion of the pellicle can be avoided by careful handling, but the difficulty arising from the variation in thickness is more serious. If a flat glass plate is carefully levelled and a layer of pyroxyline solution is poured into it and allowed to evaporate, it might be expected that the resulting film would be optically parallel, but this is not in fact the case. If the layer is thick and evaporation proceeds rapidly the surface of the pellicle is found to show an 'orange-peel' structure. With a thinner layer and very slow evaporation, which can be secured by surrounding the plate with a metal ring covered with one or more layers of filter paper, the lack of uniformity is no longer apparent to the eye, but is at once seen when examined by appropriate interference methods. A number of explanations of the 'orange-peel' effect have been suggested but the phenomenon is probably very complex.

Some years ago I noticed that if a pellicle taken from a grating was squeegeed with the rulings down, on to a plate of glass coated with gelatine, the pellicle could be stripped from the plate when the water in the gelatine had completely evaporated, and that a cast of the rulings was left on the surface of the gelatine. The advantage of this procedure was obvious. It is very much easier to prepare a flat surface of gelatine than one of pyroxyline, and the somewhat irregular pellicle of pyroxyline or cellulose acetate is discarded when once it has impressed its surface pattern on the gelatine. If the gelatine is soaked in a 4 % solution of ammonium bichromate and the pellicle is then squeegeed on to it we are left with a grating in bichromated gelatine, and if this grating is slowly heated to 200° C an enamel is formed which is remarkably hard and which can be washed with water or a mild detergent without damage and which can readily be aluminized. A further advantage of the gelatine replicas lies in the fact that they can be 'spread' with cellulose acetate and further copies can be made if the original grating is no longer available. The preparation of optically flat layers of gelatine and an investigation of the refinements in the technique which are needed for the production of substantially perfect copies of gratings is now being carried out by Dr L. A. Sayce at the National Physical Laboratory and need not be further described in the present communication.

It should be possible to rule the grating as a very fine screw thread on a cylinder, and it seemed that this procedure might simplify the construction of a ruling machine, particularly for ruling large gratings for astronomical purposes. If the pellicle gelatine process can be used successfully to prepare copies of gratings, it should be easy to coat a ruled cylinder with a pellicle of cellulose acetate and on removing this pellicle from the cylinder to lay it down on a flat surface of gelatine in exactly the same way as the pellicle from a flat grating is treated. In preparing a pellicle from a cylinder a fine wire is fastened by pieces of adhesive tape at the two ends of the cylinder so that the wire lies on the surface of the cylinder parallel to the axis. The cylinder is then 'spread' by slowly rotating it in a lathe after pouring a solution of cellulose acetate on a straight edge which is held in the tool holder. With

practice it is easy to spread uniform pellicles in this way and when they are dry it is only necessary to tear them longitudinally by pulling off the wire when they can be readily detached from the cylinder.

Experiments with a view to studying the technique were begun by ruling a spiral of 2000 lines to the inch on a stainless steel cylinder 1 in. in diameter and about a foot in length. The experiments were made with a precision lathe, and ruling was carried out by means of a diamond which produced a groove by plastic deformation of the metal, the pressure of the diamond being governed by a small weight and the diamond itself being held in a tool holder of the conventional type used for ruling. It was realized from the outset that the accuracy of a good type of precision lathe would have to be increased by a factor of 100 to 1000 if acceptable gratings were to be made in this way, but it was felt that if the validity of the technique could be established such a machine might be worth making. The greatest difficulty lay in the preparation of the cylinders which were turned, honed, lapped and polished as carefully as possible, but the results were not entirely satisfactory, and there is little doubt that the shortcomings in the results obtained are to be attributed to imperfections in the surface and figure of the cylinders. These difficulties would probably not be a serious trouble to an investigator more skilled in optical work.

Gratings with 2000 lines to the inch were made in this way, and when the diamond was correctly set, were extremely brilliant, but the definition, as might be expected, was very poor. The periodic errors can easily be seen by the following procedure. A gelatine grating on glass is prepared from a pellicle taken from the ruled cylinder. It is then cut in half, perpendicular to the rulings, and the two halves are put together face to face in register. One of the two halves is then moved, at right angles to the rulings, about one-half the pitch of the lead-screw of the lathe and then rotated through a small angle. On examining this a moiré pattern is seen. If the grating were perfect this moiré pattern should consist of straight lines running perpendicular to the rulings. With the gratings made in the manner described above the moiré pattern consisted of a very complicated wavy pattern which repeated itself with great uniformity every quarter of an inch, this being the pitch of the lead-screw.

There are a number of independent causes for the periodic errors when a spiral is ruled on a cylinder. Of these the lead-screw and its nut are probably the chief offenders, though the bearing of the lead-screw may also make a major contribution. There is also the head-stock bearing which is less important though by no means negligible, and lastly the gear wheels. If a screw and a nut are both perfect the nut should turn on the screw freely even if it fitted the screw at every point to a degree not far short of contact. Since screws and nuts are usually far from perfect it is clear that a nut can in general turn on a screw only because it fits it loosely; the better the screw and nut the closer can be the fit. There is, however, another circumstance which would allow a nut to make a very close fit with an imperfect screw and yet turn with comparative freedom; this is the case in which the threads of the nut are made of a substance sufficiently elastic to take up the periodic errors. This principle has been made use of in a device which enables a spiral to be ruled in a manner which practically eliminates periodic error, by making part of the cylinder to be ruled act as its own lead-screw and, by making the nut which fits on this part of the cylinder also carry

the diamond which rules another portion of the cylinder, all thrust bearings are automatically eliminated.

The apparatus used is shown in figure 1. It is far from perfect and both design and workmanship are in need of improvement, but it has nevertheless given surprisingly good results. Here *C* is a polished cylinder of stainless steel, one end of which is held in the collet and the other supported by the back centre of the lathe. On to this fits the nut *N* which is made in two halves which are held together by screws, one of which can be seen in the figure. Attached to the nut is a rod which terminates in the ball-race *M* which rests on the flat rod *A* which is held in the tool-rest of the lathe. The nut

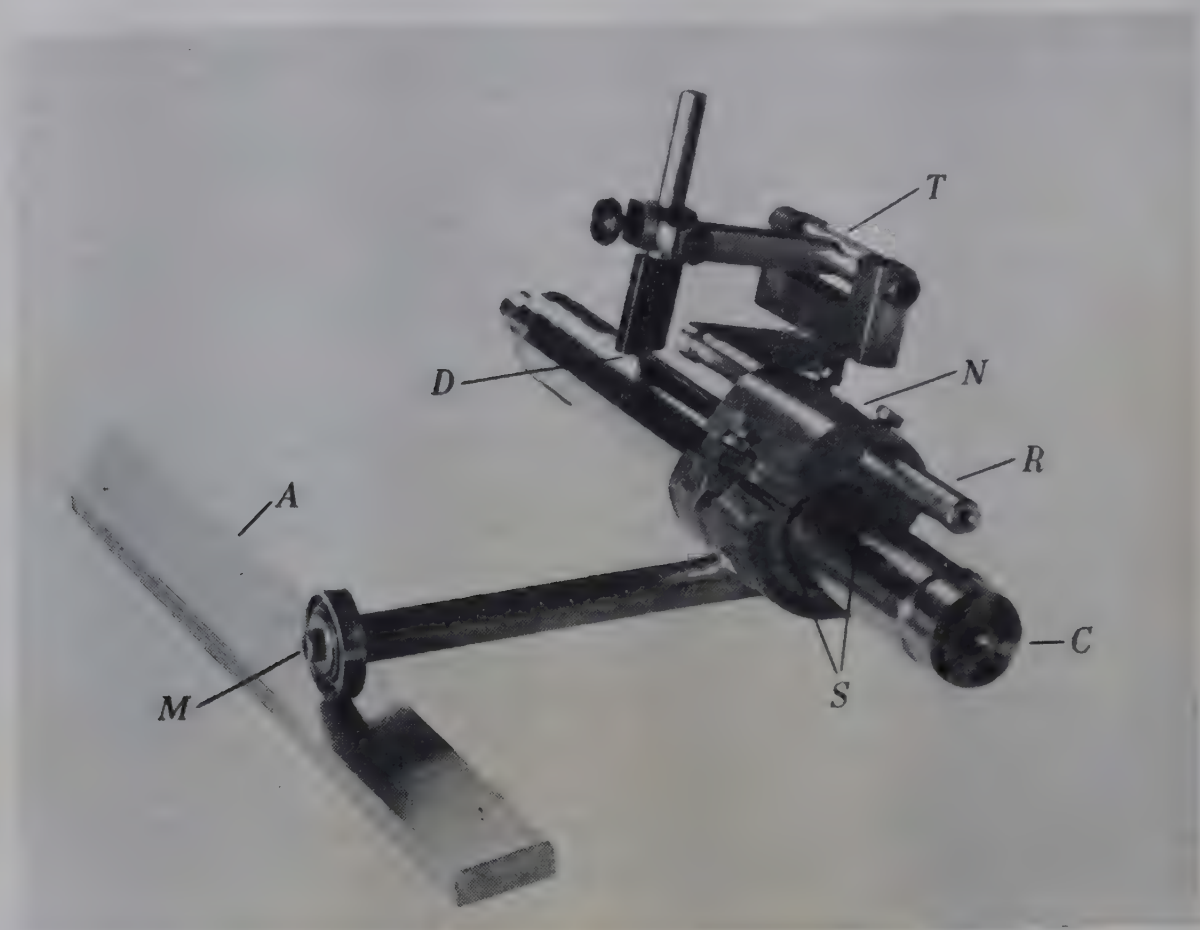


FIGURE 1

also carries an adjustable rod *R* which can be fixed in position by the two screws which can be seen at the top of the nut. This rod *R* carries the tool holder *T* which in turn carries the diamond *D*. Before the apparatus is assembled the cylinder *C* is ruled on the lathe with a spiral of 2000 turns to the inch in the usual way for a distance exceeding half the length of the cylinder by about 1 in., which is half the length of the nut. This spiral has, of course, all the periodic errors inherent in the lathe. When the apparatus is assembled the nut is fitted over the ruled portion and the diamond is adjusted over the unruled portion of the cylinder. The nut is simply a brass tube about $\frac{1}{8}$ in. larger in diameter than the cylinder, and contact is made with the cylinder by means of three slips of cork *S* about $\frac{3}{8}$ in. wide, running the whole length of the nut, the centres of these slips making an equilateral triangle when seen end-on.

The slips are cemented inside the brass nut and before use they are made true by carefully working the nut on a steel rod coated with very fine emery paper. The slips are coated with graphite and when the apparatus is properly adjusted the nut fits rather tightly on the cylinder. Other materials such as leather or even balsa wood can be used, but thin cork has given good results. When the cylinder is rotated the nut moves along one half of the cylinder and the diamond rules a spiral on the other half. The lathe is here used simply to rotate the cylinder and its accuracy is irrelevant. For example, a periodic error in the bearing of the head-stock would move both the nut and the diamond and would produce no periodic error in the ruling. Since the nut is 2 in. long it covers some 4000 turns of spiral and the cork, which probably acts in part as a nut with elastic threads and partly by slipping, averages out the periodic error. The moiré pattern obtained from a grating made from such a ruling in the manner described on p. 189 shows that periodic errors have practically disappeared, the moiré lines being nearly straight with only very slight irregularities, which can without doubt be attributed to the imperfections which were known to exist in the figure and surface of the cylinder. This procedure can be repeated as often as the length of the cylinder permits.

A grating with 2000 lines to the inch, about $2\frac{1}{2}$ in. square and substantially free from periodic error, has been made by this method, but the results must be considered as preliminary. In particular, the choice of a suitable metal for the cylinders may call for much experiment. A modification of the design to enable gratings to be ruled on flat or concave surfaces, without violating the conditions which are essential to the elimination of periodic errors, follows on fairly obvious lines, which it is hoped to pursue. It is hoped that these experiments may be of use in other fields where an exceedingly slow motion with freedom from periodic error, is desired.

I am indebted to Dr H. H. Burton for kindly supplying me with some rods of stainless steel, and to Mr F. Morris for his assistance in these experiments.

REFERENCE

Wallace, R. J. 1905 *Astrophys. J.* **22**, 123.

The instability of liquid surfaces when accelerated in a direction perpendicular to their planes. I

BY SIR GEOFFREY TAYLOR, F.R.S.

(Received 3 December 1949)

It is shown that, when two superposed fluids of different densities are accelerated in a direction perpendicular to their interface, this surface is stable or unstable according to whether the acceleration is directed from the heavier to the lighter fluid or vice versa. The relationship between the rate of development of the instability and the length of wave-like disturbances, the acceleration and the densities is found, and similar calculations are made for the case when a sheet of liquid of uniform depth is accelerated.

1. INTRODUCTION

If the horizontal surface of a liquid at rest under gravity is displaced into the form of regular small corrugations and then released, standing oscillatory waves are produced. Theoretically, a liquid could exist in a state of unstable equilibrium with flat lower horizontal surface supported by air pressure. If small wave-like corrugations were formed on this lower surface and then released, they might be expected to increase exponentially so long as their height was small compared with the wave-length.

This instability of the lower surface of a liquid might be expected to disappear if the liquid were allowed to fall freely, and to pass over into stability if the liquid were forced downwards with an acceleration greater than that of gravity. Similarly, the initial stability of the upper surface of a liquid might be expected to pass over into instability if the liquid were given a downward acceleration greater than that of gravity.

2. THE INSTABILITY OF AN INTERFACE BETWEEN TWO FLUIDS

Consider the interface between two fluids of densities ρ_1 and ρ_2 . Assuming that they are accelerated vertically upwards with acceleration g_1 , take horizontal and vertical axes which are fixed relative to the interface. The equations of motion relative to the accelerating axes are identical with those of a fluid at rest except that a term $-g_1\rho_1y$ must be added to the pressure in the upper fluid and a term $-g_1\rho_2y$ to that in the lower fluid. If both fluids are deep compared with the wave-length of the disturbance of the interface, the velocity potentials of the motion are

$$\phi_1 = Ae^{-Ky+nt} \cos Kx \quad \text{in the upper fluid,} \quad (1)$$

$$\phi_2 = -Ae^{Ky+nt} \cos Kx \quad \text{in the lower fluid.} \quad (2)$$

The equation of the surface of separation is then

$$\eta = AKn^{-1}e^{nt} \cos Kx. \quad (3)$$

The pressure in the upper fluid, neglecting squares of disturbance velocity, is

$$p_1 = p - (g + g_1)\rho_1y + \rho_1\phi_1,$$

where p is the mean pressure at the interface and g is the acceleration of gravity.

Similarly, the pressure in the lower fluid is

$$p_2 = p - (g + g_1) \rho_2 y + \rho_2 \cdot \phi_2.$$

At the interface, therefore

$$-(g + g_1) (\rho_2 - \rho_1) y = (\rho_1 + \rho_2) n A e^{nt} \cos Kx. \quad (4)$$

Hence from equation (3)

$$n^2 = -K(g + g_1) \frac{(\rho_2 - \rho_1)}{(\rho_2 + \rho_1)}. \quad (5)$$

If $(g + g_1)$ is negative, the acceleration downwards is greater than g and n^2 is positive. In this case the velocity potentials are of the form

$$\phi_1 = (A e^{nt - Ky} + B e^{-nt - Ky}) \cos Kx, \quad (6)$$

$$\phi_2 = -(A e^{nt + Ky} + B e^{-nt + Ky}) \cos Kx, \quad (7)$$

where n is taken to be the positive value of

$$\sqrt{\left(-K(g + g_1) \frac{(\rho_2 - \rho_1)}{(\rho_2 + \rho_1)}\right)}.$$

If the initial displacement of the interface from the plane $y = 0$ is

$$\eta_0 = C \cos Kx,$$

and the initial velocity is zero, equations (3), (6) and (7) become

$$\eta = C \cosh nt \cos Kx, \quad (8)$$

$$\phi_1 = \frac{nC}{K} e^{-Ky} \sinh nt \cos Kx, \quad (9)$$

$$\phi_2 = -\frac{nC}{K} e^{Ky} \sinh nt \cos Kx. \quad (10)$$

The initial disturbance of the interface will increase exponentially with nt until it has attained a magnitude which is no longer small in comparison with the wavelength. The rate of development of the instability is proportional to $\sqrt{\frac{(\rho_2 - \rho_1)}{(\rho_2 + \rho_1)}}$.

3. THE INSTABILITY OF A LIQUID SHEET OF FINITE THICKNESS h ACCELERATED BY AIR PRESSURE

In this case it will be assumed that the liquid sheet is of uniform thickness and is accelerated by air pressure acting on its upper surface. If the density of the liquid is great compared with that of air, the pressure may be assumed uniform over both the upper and lower surfaces.

Assuming in the first place that the elevations of the disturbances above their mean levels are of the form

$$\eta_U = \alpha e^{nt} \cos Kx \quad \text{at the upper surface} \quad (11)$$

and $\eta_L = \beta e^{nt} \cos Kx$ at the lower surface, (12)

these are consistent with a velocity potential

$$\phi = (Ae^{Ky} + Be^{-Ky}) e^{nt} \cos Kx, \quad (13)$$

provided that $\frac{\partial \eta_U}{\partial t} = -\frac{\partial \phi}{\partial y}$ at $y = 0$,

and $\frac{\partial \eta_L}{\partial t} = -\frac{\partial \phi}{\partial y}$ at $y = -h$,

so that $\alpha n = -K(A - B)$, (14)

and $\beta n = -K(Ae^{-Kh} + Be^{Kh})$. (15)

The pressure in the fluid (neglecting squares of disturbance velocity) is

$$p = P - \rho y(g + g_1) + \rho \phi,$$

where P is the pressure at the upper surface. The condition that p is constant at $y = \eta_U$ is

$$\alpha(g + g_1) = n(A + B), \quad (16)$$

while the condition at $y = \eta_L$ is

$$\beta(g + g_1) = n(Ae^{-Kh} + Be^{Kh}). \quad (17)$$

The four equations (14), (15), (16) and (17) can only be satisfied simultaneously if either $A = 0$ or $B = 0$.

If $B = 0$, then $\alpha = \beta e^{Kh}$, $\alpha n = -KA$ and $\alpha(g + g_1) = nA$,

so that $n^2 = -K(g + g_1)$. (18)

If $A = 0$, then $\alpha = \beta e^{-Kh}$, $\alpha n = KB$ and $\alpha(g + g_1) = nB$,

so that in this case $n^2 = K(g + g_1)$. (19)

By combining together solutions of these two types, the initial conditions at the top and bottom surfaces can be satisfied. As an example, take the case when the liquid is initially at rest but with the upper and lower surfaces corrugated into waves of amplitudes α_0 and β_0 respectively. Supposing that the liquid is accelerated downwards by applying a pressure at the upper surface, $(g + g_1)$ is then negative and the appropriate form for the velocity potential is

$$\phi = (Ae^{Ky} \sinh nt + Be^{-Ky} \sin nt) \cos Kx, \quad (20)$$

where n is the positive value of $\sqrt{\{ -K(g + g_1) \}}$.

The elevation of the upper and lower surfaces are

$$\eta_U = \frac{K}{n} (-A \cosh nt - B \cos nt) \cos Kx, \quad (21)$$

and $\eta_L = \frac{K}{n} (Ae^{-Kh} \cosh nt - Be^{Kh} \cos nt) \cos Kx$. (22)

When $t = 0$, $\eta_U = \alpha_0 \cos Kx$ and $\eta_L = \beta_0 \cos Kx$.

Therefore

$$A + B = -\alpha_0 \frac{n}{K}, \quad (23)$$

and

$$Ae^{-Kh} + Be^{Kh} = -\beta_0 \frac{n}{K}. \quad (24)$$

Solving for A and B and substituting in equations (21) and (22), the elevations of the upper and lower surfaces become

$$\eta_U = \{(\alpha_0 - \beta_0 e^{-Kh}) \cosh nt - (\alpha_0 e^{-2Kh} - \beta_0 e^{-Kh}) \cos nt\} \frac{\cos Kx}{1 - e^{-2Kh}}, \quad (25)$$

$$\eta_L = \{(\alpha_0 e^{-Kh} - \beta_0 e^{-2Kh}) \cosh nt - (\alpha_0 e^{-Kh} - \beta_0) \cos nt\} \frac{\cos Kx}{1 - e^{-2Kh}}. \quad (26)$$

In the experimental work described by Mr D. J. Lewis in part II (1950), various liquids were accelerated with an initial disturbance on the upper surface and with the lower surface initially flat. This corresponds to the case of $\beta_0 = 0$, when equations (25) and (26) reduce to

$$\eta_U = \alpha_0 \left\{ \frac{\cosh nt - e^{-2Kh} \cos nt}{1 - e^{-2Kh}} \right\} \cos Kx, \quad (27)$$

$$\eta_L = \alpha_0 \left\{ \frac{\cosh nt - \cos nt}{1 - e^{-2Kh}} \right\} e^{-Kh} \cos Kx. \quad (28)$$

The effect of the lower surface on the upper is represented by the terms in equation (27) containing e^{-2Kh} as a factor. If h is greater than $\frac{1}{3}\lambda$, where λ is the wave-length, e^{-2Kh} is less than 0.015. Neglecting the terms containing e^{-2Kh} , equation (27) becomes

$$\eta_U = \alpha_0 \cosh nt \cos Kx, \quad (29)$$

which is the form appropriate to a liquid of infinite depth.

In the experiments described in part II this approximation seems to be justifiable.

4. CONCLUSIONS

Defining the amplification factor of an unstable fluid surface as the ratio of the amplitude of the disturbance at any time to its initial value, we have for a vertical downward acceleration g_1

$$\frac{\eta}{\eta_0} = \cosh nt = \cosh \left\{ -K(g_1 - g) \frac{(\rho_2 - \rho_1)}{(\rho_2 + \rho_1)} \right\}^{\frac{1}{2}} t. \quad (30)$$

This applies to an initially disturbed interface between two fluids and also approximates closely to the behaviour of the upper surface of a sheet of liquid of density ρ_2 of thickness $h > \frac{1}{3}\lambda$ when accelerated by air pressure, in which case ρ_1 may be neglected by comparison with ρ_2 . The acceleration is taken to begin at time $t = 0$, and at time t the liquid has moved downwards a distance s given by

$$s = \frac{1}{2} g_1 t^2,$$

so that

$$\frac{\eta}{\eta_0} = \cosh \sqrt{\left\{ \frac{4\pi s}{\lambda} \frac{(g_1 - g)}{g_1} \frac{(\rho_2 - \rho_1)}{(\rho_2 + \rho_1)} \right\}} t. \quad (31)$$

If g_1 is large compared with g ,

$$\frac{\eta}{\eta_0} = \cosh \sqrt{\left\{ \frac{4\pi s}{\lambda} \frac{(\rho_2 - \rho_1)}{(\rho_2 + \rho_1)} \right\}} = \cosh \sqrt{\left\{ 4\pi m \frac{(\rho_2 - \rho_1)}{(\rho_2 + \rho_1)} \right\}}, \quad (32)$$

where m is the number of wave-lengths that the liquid has descended. Thus, when $g_1 \gg g$, the amplification of an unstable disturbance depends only on the ratio of the densities of the two fluids and the number of wave-lengths through which they have descended.

The experiments described in part II were designed to test the above theoretical conclusions and to find out how the unstable surface behaves when the amplitude of the disturbance becomes so great that the analysis based on the assumption that it is small ceases to be applicable.

REFERENCE

Lewis, D. J. 1950 *Proc. Roy. Soc. A* (in the Press).

The electronic structure of conjugated systems. VI

BY B. H. CHIRGWIN, *Queen Mary College, London*

AND C. A. COULSON, *Wheatstone Physics Laboratory, King's College, London*

(Communicated by J. T. Randall, F.R.S.—Received 18 August 1949)

This part reconsiders the previously neglected non-orthogonality of atomic orbitals. New generalized definitions of fractional bond order and charge distribution are introduced, and it is shown that under certain conditions, which are satisfied by most conjugated hydrocarbons, the overlap factor is completely irrelevant in calculating bond orders and charges. For heteromolecules this is not the case, and substantial changes may sometimes be made by its inclusion. The various polarizability coefficients introduced in earlier papers are always affected, but numerical results suggest that these changes are not very large.

INTRODUCTION

Previous papers in this series (Coulson & Longuet-Higgins 1947–8, referred to as I to V) have discussed in quite general terms the bond orders and charge distribution in conjugated systems; and have introduced various polarizability coefficients whose values showed how a molecule was likely to be affected by a change in the environment of any part of it. Applications were made to the law of alternating polarity in chemical reactions, to the vibrations of condensed molecules such as benzene and naphthalene, to the conjugating power of unsaturated free radicals, and to the differences between hydrocarbon systems such as naphthalene and heterosystems such as quinoline, in which a CH group has been replaced by a N atom. But in all this work it was explicitly supposed that all overlap integrals $S_{\mu\nu} = \int \phi_\mu \phi_\nu d\tau$ between distinct atomic orbitals ϕ_μ and ϕ_ν were identically zero. Such an approximation has almost universally been made hitherto, despite the fact that numerical calculations

show that for two adjacent carbon π -electron orbitals $S_{\mu\nu}$ is in the region of 0.25. This is very far from being negligible in comparison with the $S_{\mu\mu}$, which are all unity, since each ϕ_μ is normalized. The only general discussions of the possibility of including $S_{\mu\nu}$ of which we are aware are for hydrocarbons by Wheland (1941), who showed how to correct the energy values previously obtained by neglect of $S_{\mu\nu}$, and for graphitic crystallites by Bradburn, Coulson & Rushbrooke (1948), who were also chiefly concerned with energy values. Both of these papers dealt with only one kind of unsaturated atom, here carbon. It is our purpose in this paper to discuss quite generally the significance of $S_{\mu\nu}$ for molecules in which the unsaturated atoms are not all of one kind. This will involve us in new definitions of bond order and charge density. Under certain circumstances, of which the contents of the papers referred to above are particular examples, it will be shown that inclusion of $S_{\mu\nu}$ makes no difference whatever to charge density or bond orders, though it does always make numerical changes in the polarizability coefficients. However, these latter changes appear often to be small. But with heteromolecules, both charges and bond orders are sensitive to the values of the $S_{\mu\nu}$.

NOTATION

So far as possible we shall use the notation of previous papers of this series. But certain changes are necessitated by our extension of earlier theorems. In particular, we shall denote different atoms by a Greek suffix $\mu, \nu, \rho, \dots$ and a molecular orbital (m.o.) by a Latin suffix $i, j, \dots$. In the usual LCAO approximation (see earlier papers in this series), we expand each m.o. ψ_j in the form

$$\psi_j = \sum_{\mu=1}^n x_{j\mu} \phi_\mu. \quad (1)$$

When there is no doubt about which m.o. we are discussing, we can drop the suffix j and write simply x_μ instead of $x_{j\mu}$. There are n atoms in the molecule, and the coefficients $x_{j\mu}$ are determined by minimizing the expression for the energy

$$E = \int \psi^* H \psi d\tau / \int \psi^* \psi d\tau. \quad (2)$$

This leads to the set of n equations (the secular equations)

$$\sum_{\nu=1}^n (H_{\mu\nu} - ES_{\mu\nu}) x_\nu = 0 \quad (\mu = 1, 2, \dots, n), \quad (3)$$

and thus to the secular determinant

$$\det | H_{\mu\nu} - ES_{\mu\nu} | = 0. \quad (4)$$

Solution of this determinant provides us with the energies E_j ($j = 1, 2, \dots, n$) of the possible orbitals, and (3) gives the corresponding coefficients $x_{j\mu}$. If the m.o. ψ_j is normalized, so that

$$\int \psi_j^2 d\tau = 1, \quad (5)$$

then

$$\sum_{\mu, \nu=1}^n S_{\mu\nu} x_{j\mu} x_{j\nu} = 1. \quad (6)$$

It will be noticed that we have used $x_{j\mu}$ for the coefficients in the m.o. (1). This allows us to reserve the previously used $c_{j\mu}$ for the earlier approximations, in which $S_{\mu\nu} = 0$. As we shall see, there are important relations between $x_{j\mu}$ and $c_{j\mu}$. For the same reason let us use E , or E_j , for the energy found by solution of the full secular determinant (4), and ϵ , or ϵ_j , for the corresponding energy as calculated with neglect of $S_{\mu\nu}$.

NEW DEFINITIONS OF BOND ORDER AND CHARGE DENSITY

We must now proceed to new definitions of bond order and charge density. The old definitions, according to which the mobile bond order $p_{\mu\nu}$ between adjacent atoms μ, ν and the total π -charge q_μ on atom μ , were given by

$$p_{\mu\nu} = \sum_{j=1}^m n_j c_{j\mu} c_{j\nu}, \quad (7)$$

$$q_\mu = \sum_{j=1}^m n_j c_{j\mu}^2, \quad (8)$$

the summation being taken over all the allowed m.o., and n_j being the number of electrons in the j th orbital, will no longer serve. In the first place, the normalization condition (6) shows that $\sum_{\mu} q_\mu$ would no longer be equal to n , so that if we retained (8) with $x_{j\mu}$ replacing $c_{j\mu}$, we should be leaving out fractional parts of each electron; and in the second place, for the same reason, p and q would no longer satisfy the same simple differential relations (I, equations (24) and (25))

$$p_{\mu\nu} = \frac{1}{2} \frac{\partial \mathcal{E}}{\partial \beta_{\mu\nu}}, \quad q_\mu = \frac{\partial \mathcal{E}}{\partial \alpha_\mu}.$$

The existence of these relations was quite fundamental to the whole subsequent development of the various self and mutual polarizabilities.

To obtain new definitions we must return to (6), which shows that $S_{\mu\nu}$ may be regarded as the components of the metric tensor of a space defined by the x_μ . This suggests that we should introduce a new set of coefficients $y_{j\mu}$ associated with each ψ_j (or $x_{j\mu}$), and defined by the relations

$$y_{j\mu} = \sum_{\nu=1}^n S_{\mu\nu} x_{j\nu} \quad \text{or} \quad x_{j\mu} = \sum_{\nu=1}^n S^{\mu\nu} y_{j\nu}. \quad (9)$$

We have taken $S^{\mu\nu}$ to be the components of the tensor reciprocal to $S_{\mu\nu}$, so that

$$\left. \begin{aligned} \sum_{\nu=1}^n S^{\lambda\nu} S_{\mu\nu} &= \delta_\mu^\lambda = 1 \quad (\lambda = \mu), \\ &= 0 \quad (\lambda \neq \mu). \end{aligned} \right\} \quad (10)$$

On account of (9) we may call the $y_{j\mu}$ the 'covariant' coefficients associated with ψ_j ; evidently $x_{j\mu}$ are the 'contravariant' coefficients. The introduction of the $y_{j\mu}$ coefficients allows us to write the normalization equation (6) in the more suggestive form

$$\sum_{\nu=1}^n x_{j\nu} y_{j\nu} = 1 \quad (j = 1, 2, \dots, n). \quad (11)$$

In the same way, since the ψ_j correspond to principal directions of the Hamiltonian matrix $H_{\mu\nu}$ ($H_{\mu\nu} = \int \phi_\mu H \phi_\nu d\tau$), the different ψ_j are mutually orthogonal, so that, as may easily be shown directly from the secular equations

$$\sum_{\nu=1}^n x_{i\nu} y_{j\nu} = 0 \quad (i \neq j). \quad (12)$$

We now postulate that the former definitions (7) and (8) of bond order and charge are to be replaced by

$$q_\mu = \sum_{j=1}^m n_j x_{j\mu} y_{j\mu} \quad (\mu = 1, 2, \dots, n), \quad (13)$$

$$p_{\mu\nu} = \frac{1}{2} \sum_{j=1}^m n_j (x_{j\mu} y_{j\nu} + x_{j\nu} y_{j\mu}). \quad (14)$$

In (13) and (14) the summation covers all the occupied m.o. ($j = 1, \dots, m$) and n_j is the number of electrons in the j th orbital.

It is easily verified that when the non-diagonal $S_{\mu\nu}$ are put equal to zero, as in earlier treatments, (13) and (14) reduce to (7) and (8). For, as (9) shows, the distinction between $x_{j\mu}$ and $y_{j\mu}$ disappears, and both reduce to the $c_{j\mu}$ involved in (7) and (8). These $c_{j\mu}$ satisfy the equations (cf. (3) and (6))

$$\sum_{\nu=1}^n (H_{\mu\nu} - \epsilon_j \delta_{\mu\nu}) c_{j\nu} = 0, \quad (15)$$

$$\sum_{\nu=1}^n c_{j\nu}^2 = 1. \quad (16)$$

Equations (15) and (16) show that the new definitions of q_μ and $p_{\mu\nu}$ are indeed generalizations of the old. We shall show later, in (32) to (34) and (39) to (41), that they are satisfactory generalizations because they are successful in preserving the fundamental differential relationships between the total energy and the matrix elements $H_{\mu\nu}$.

CONDITIONS FOR EQUIVALENCE OF OLD AND NEW BOND ORDERS AND CHARGES

Calculations in several special cases, e.g. benzene, butadiene, show that when the overlap integral $S_{\mu\nu}$ is given a constant value S for all nearest neighbours, then the contributions to the bond orders and charge densities from a given m.o. have a value independent of S , and hence the same as in the earlier forms of the theory in which S was completely neglected. This is a special case of a more general result which may be stated in the form that when the matrices $H_{\mu\nu}$ and $S_{\mu\nu}$ commute, the bond orders and charges computed according to (13) and (14) with inclusion of overlap are the same as those computed according to (7) and (8) with neglect of overlap.

First let us show that this condition is necessary. If we suppose that (13) and (14) give the same numerical values as (7) and (8), then

$$x_{j\mu} y_{j\mu} = c_{j\mu}^2, \quad x_{j\mu} y_{j\nu} + x_{j\nu} y_{j\mu} = 2c_{j\mu} c_{j\nu}.$$

From the first of these equations it follows that we may write

$$x_{j\mu}/c_{j\mu} = c_{j\mu}/y_{j\mu} = \lambda_{j\mu}, \quad (17)$$

and now, from the second equation,

$$\lambda_{j\mu}/\lambda_{j\nu} + \lambda_{j\nu}/\lambda_{j\mu} = 2.$$

Hence

$$\lambda_{j\mu} = \lambda_{j\nu} \quad (= \lambda_j, \text{ say}). \quad (18)$$

Thus

$$x_{j\mu} = \lambda_j c_{j\mu}, \quad y_{j\mu} = c_{j\mu}/\lambda_j \quad \text{and} \quad x_{j\mu} = \lambda_j^2 y_{j\mu}. \quad (19)$$

But from the definition of the $y_{j\mu}$ in (9) we have

$$\sum_{\nu=1}^n S_{\mu\nu} \lambda_j c_{j\nu} = c_{j\mu}/\lambda_j,$$

i.e.

$$\sum_{\nu=1}^n \left(S_{\mu\nu} - \frac{1}{\lambda_j^2} \delta_{\mu\nu} \right) c_{j\nu} = 0. \quad (20)$$

This shows that the $c_{j\mu}$ give the principal directions of the tensor $S_{\mu\nu}$. Equation (15) shows that this is also true for $H_{\mu\nu}$. It follows that **S** and **H** commute, as stated to be necessary in the theorem.

This analysis also provides a simple relation between the energy ϵ_j calculated without overlap, and E_j calculated with overlap. For, by (3),

$$\sum_{\nu} H_{\mu\nu} x_{j\nu} = \sum_{\nu} E_j S_{\mu\nu} x_{j\nu},$$

and so, using (9) and (19),

$$\lambda_j \sum_{\nu} H_{\mu\nu} c_{j\nu} = (E_j/\lambda_j) c_{j\mu}.$$

Comparison with (15) shows that

$$\epsilon_j = E_j/\lambda_j^2. \quad (21)$$

We have shown above that a necessary condition that the contributions to bond orders and charge densities from each separate orbital should be the same in the two treatments is that $H_{\mu\nu}$ and $S_{\mu\nu}$ have the same principal directions and hence should commute. It may be shown as follows that the condition is sufficient.

For, if the two tensors have the same principal directions $c_{j\mu}$, then

$$\sum_{\nu} H_{\mu\nu} c_{j\nu} = \epsilon_j c_{j\mu}, \quad \sum_{\nu} S_{\mu\nu} c_{j\nu} = (1/\lambda_j^2) c_{j\mu}.$$

Hence, if E_j is chosen to be equal to $\lambda_j^2 \epsilon_j$, it follows that

$$\sum_{\nu} (H_{\mu\nu} - E_j S_{\mu\nu}) c_{j\nu} = (\epsilon_j - E_j/\lambda_j^2) c_{j\mu} = 0.$$

Thus E_j is the energy when overlap is included, so that the relation (21) is reproduced: and also the $x_{j\mu}$ and $c_{j\mu}$ must be proportional. If we write

$$x_{j\mu} = k c_{j\mu},$$

then the normalizing condition (6) gives that

$$\begin{aligned} 1 &= \sum_{\mu, \nu} S_{\mu\nu} x_{j\mu} x_{j\nu} = k^2 \sum_{\mu, \nu} S_{\mu\nu} c_{j\mu} c_{j\nu} \\ &= k^2 \sum_{\mu} c_{j\mu}^2 / \lambda_j^2 \\ &= k^2 / \lambda_j^2, \end{aligned}$$

if we suppose that the $c_{j\mu}$ are normalized according to (16). From this it follows that $k = \lambda_j$, which gives (19) and thus equal contributions to charges and bond orders in both old and new theories. This establishes both the necessary and sufficient character of the condition stated at the beginning of this section.

SOME PARTICULAR CASES

Before proceeding further, it will be convenient to discuss some particular cases of the applicability or otherwise of the preceding analysis.

(i) If all non-diagonal $S_{\mu\nu}$ are zero, as in the usual development of the theory, then $\mathbf{S}$ is a unit matrix, and quite evidently $\mathbf{S}$ and $\mathbf{H}$ commute, whatever the form of $H_{\mu\nu}$.

(ii) In hydrocarbon molecules it is usual to suppose that all the diagonal terms $H_{\mu\mu}$ are equal, with a value $\alpha (= H_{CC})$, and that all $H_{\mu\nu}$ are zero except when μ and ν are neighbouring atoms, in which case $H_{\mu\nu} = \beta$, for all μ, ν . If we make the corresponding assumption that all $S_{\mu\nu}$ are zero except between neighbours and that then $S_{\mu\nu} = S$ for all μ, ν , it is easily verified that $\mathbf{H}$ and $\mathbf{S}$ commute. This shows that the bond orders calculated for these molecules with neglect of S are entirely unaffected, though as Wheland (1941) has shown, this is not true for the energies.

This particular case is probably the most important of the applications of our theory, and is worth pursuing further. Thus a particular equation of the set (3) is

$$\beta x_{j\pi} + \alpha x_{j\rho} + \beta x_{j\sigma} + \beta x_{j\tau} = E_j y_{j\rho},$$

where π, σ, τ are the neighbouring atoms to ρ . The corresponding equation (9) relating $x_{j\mu}$ and $y_{j\mu}$ is

$$Sx_{j\pi} + x_{j\rho} + Sx_{j\sigma} + Sx_{j\tau} = y_{j\rho}.$$

If we eliminate $x_{j\pi}$ between these two equations, we have

$$(\beta - \alpha S)x_{j\rho} = (\beta - E_j S)y_{j\rho},$$

so that, as in the third equation of (19),

$$x_{j\rho} = \lambda_j^2 y_{j\rho}, \quad (22)$$

where

$$\lambda_j^2 = \frac{\beta - E_j S}{\beta - \alpha S}. \quad (23)$$

A comparison of (21) and (23) now shows that

$$E_j = \frac{\beta \epsilon_j}{\beta + S(\epsilon_j - \alpha)}. \quad (24)$$

Now when $S = 0$, it is well known that the energies ϵ_j are always of the form

$$\epsilon_j = \alpha + m_j \beta, \quad (25)$$

where m_j is some number between ± 3 . It follows that

$$E_j = \frac{\alpha + m_j \beta}{1 + m_j S} = \alpha + \frac{m_j(\beta - \alpha S)}{1 + m_j S}, \quad (26)$$

and

$$\lambda_j = (1 + m_j S)^{-1/2}. \quad (27)$$

Equation (26) was obtained quite differently by Wheland (1941).

(iii) An immediate extension of this latter analysis is to suppose that additional values of $S_{\mu\nu}$ are to be included, so that we consider overlap between atoms other than nearest neighbours. If we suppose that the corresponding $H_{\mu\nu}$ are also included, and that the ratio $H_{\mu\nu} : S_{\mu\nu}$ is the same for all μ, ν (except, of course, when $\mu = \nu$, for which $H_{\mu\mu} = \alpha$, $S_{\mu\mu} = 1$), then $\mathbf{H}$ and $\mathbf{S}$ still commute, for it is soon recognized that if S and β are the numerical values for nearest neighbours,

$$S\mathbf{H} = \beta\mathbf{S} + (S\alpha - \beta)\mathbf{1},$$

where $\mathbf{1}$ is the unit matrix. Charge distributions and bond orders are once again independent of the value of S ; and equations (22) to (27) apply without alteration. Now Mulliken (1948) has given grounds for believing that the ratio $H_{\mu\nu} : S_{\mu\nu}$ is approximately constant for all pairs of atoms and all internuclear distances. The importance of such a fact for our theory is at once apparent.

(iv) The conclusions in (ii) and (iii) above are rendered invalid as soon as any Coulomb term for one atom differs from that for the others. This means that when we are making calculations for heteromolecules, we shall not expect to get valid results unless we include the overlap integral right from the start. Numerical calculation in one or two simple cases suggests that the resulting changes in bond orders are sometimes not very great, so that, for example, the molecular diagrams of pyridine, acridine, carbazole, etc., displayed by Longuet-Higgins & Coulson (1947) though not strictly correct, are unlikely to be much in error.

INTEGRAL FORMULAE FOR CHARGE AND BOND ORDERS

In paper I of this series (equations (42) and (43)) integral formulae were obtained for bond orders and charge densities. These formulae were especially useful in the subsequent discussion of mutual and self-polarizabilities. It becomes important to know whether any such formulae are available in the more general theory here being discussed. We shall now show that these do exist, though in order to obtain them, we are obliged to introduce 'mixed' components of the tensor $H_{\mu\nu}$, defined by

$$H_{\nu}^{\mu} = \sum_{\rho=1}^n S^{\mu\rho} H_{\nu\rho}. \quad (28)$$

These new quantities are not symmetrical in μ and ν , as the $H_{\mu\nu}$ were. But if $\mathbf{H}$ and $\mathbf{S}$ commute, then H_{μ}^{ν} does equal H_{ν}^{μ} . In the case where overlap integrals are neglected, $H_{\mu}^{\nu} = H_{\mu\nu}$.

In terms of these mixed components, the secular determinant (4) may be written

$$\det | H_{\mu}^{\nu} - E\delta_{\mu}^{\nu} | = 0, \quad (29)$$

in which the energy E appears only along the principal diagonal. This determinant is not in general symmetrical about its leading diagonal. Similarly, the secular equations (3) may be written in either of the forms

$$\sum_{\nu=1}^n (H_{\nu}^{\mu} - E\delta_{\nu}^{\mu}) x_{\nu} = 0, \quad \sum_{\mu=1}^n (H_{\nu}^{\mu} - E\delta_{\nu}^{\mu}) y_{\mu} = 0. \quad (30)$$

Equations (29) and (30) are similar to those used in I, so that our subsequent analysis follows formally along exactly the same lines. Thus we write

$$E = \sum_{\mu, \nu} (y_{\mu} H_{\nu}^{\mu} x_{\nu}) / \sum_{\mu} (x_{\mu} y_{\mu}). \quad (31)$$

If a small variation is now made in the H_{ν}^{μ} (treated as all independent), and in the x_{μ} , the variation of E is

$$\delta E = \sum_{\mu, \nu} (y_{\mu} \delta H_{\nu}^{\mu} x_{\nu}) / \sum_{\mu} (x_{\mu} y_{\mu}) + \sum_{\rho} \frac{\partial E}{\partial x_{\rho}} \delta x_{\rho}.$$

Any variations in y_{μ} due to variations in x_{μ} are included in the second term on the right-hand side. The stationary condition imposed on E to give the secular equations (30) shows that $\partial E / \partial x_{\rho} = 0$, so that

$$\left(\frac{\partial E}{\partial H_{\nu}^{\mu}} \right)_{\Delta=0} = x_{\nu} y_{\mu}. \quad (32)$$

Summation over the occupied orbitals gives

$$q_{\mu} = \sum_{j=1}^m n_j \left(\frac{\partial E_j}{\partial H_{\mu}^{\mu}} \right)_{\Delta=0} = \left(\frac{\partial \mathcal{E}}{\partial H_{\mu}^{\mu}} \right)_{\Delta=0}, \quad (33)$$

$$p_{\mu\nu} = \frac{1}{2} \sum_{j=1}^m n_j \left[\frac{\partial E_j}{\partial H_{\mu}^{\nu}} + \frac{\partial E_j}{\partial H_{\nu}^{\mu}} \right]_{\Delta=0} = \frac{1}{2} \left(\frac{\partial \mathcal{E}}{\partial H_{\mu}^{\nu}} + \frac{\partial \mathcal{E}}{\partial H_{\nu}^{\mu}} \right)_{\Delta=0}. \quad (34)$$

We have written $\mathcal{E}$ for the total energy of the mobile electrons, defined by

$$\mathcal{E} = \sum_{j=1}^m n_j E_j, \quad (35)$$

and the notation $()_{\Delta=0}$ means, as in I, that the differentiations involved must always be made under the condition that the various variables satisfy the secular determinant $\Delta = 0$ in (29).

The expressions (33) and (34) may be converted into contour integrals as in I, pp. 46-49. The result is

$$q_{\mu} = 1 - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\partial}{\partial H_{\mu}^{\mu}} \log \Delta(iy) dy, \quad (36)$$

$$\begin{aligned} p_{\mu\nu} &= (-1)^{\mu+\nu+1} \frac{1}{2\pi} \int \frac{\Delta_{\mu,\nu}(iy) + \Delta_{\nu,\mu}(iy)}{\Delta(iy)} dy \\ &= -\frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial H_{\mu}^{\nu}} + \frac{\partial}{\partial H_{\nu}^{\mu}} \right) \log \Delta(iy) dy. \end{aligned} \quad (37)$$

There is a similar integral relation for $\mathcal{E}$. Thus (cf. I (44))

$$\mathcal{E} = \sum H_{\mu}^{\mu} + \frac{1}{\pi} \int_{-\infty}^{\infty} \left\{ iy \frac{\Delta'(iy)}{\Delta(iy)} - n \right\} dy. \quad (38)$$

In these expressions $\Delta(iy)$ stands for the unsymmetrical determinant of (29) with iy replacing E ; $\Delta_{\mu,\nu}(iy)$ is the same determinant with row μ and column ν struck out. In (36) to (38) we have tacitly assumed that the occupied orbitals are all bonding, and the unoccupied are anti-bonding, so that $E = 0$ lies between E_m and E_{m+1} . This is nearly always the case, but may not always be so if some of the heteroatoms present have

very large Coulomb terms. If this should happen, the determinants in these equations should be regarded as functions of $x + iy$, rather than iy , where x is any value between E_m and E_{m+1} . Provided that x lies in this range, all three integrals are actually independent of its precise value. Formulae (36) to (38) are similar to (42) to (44) of paper I, and do indeed reduce to them when overlap integrals are neglected. The most significant difference is that our new formulae involve the components H_μ^ν of the mixed tensor rather than the components $H_{\mu\nu}$ of the original tensor. This suggests very strongly that it is these mixed components H_μ^μ and H_μ^ν that have physical significance rather than the simple Coulomb integrals $H_{\mu\mu}$ and resonance integrals $H_{\mu\nu}$. Now the essential difference between H_μ^μ and $H_{\mu\mu}$ is that whereas the latter is a function only of atom μ , the former depends not only on μ but also on all the neighbours of μ . In this respect, then, our new interpretation is likely to be better than the old one, for as was explained in I, p. 43, we expect the diagonal elements $H_{\mu\mu}$ (or H_μ^μ) to be a measure of the electron affinity of atom μ ; and it is at least plausible that this electron affinity should depend not only on μ , but also, to a lesser extent, on the neighbours of μ . Some recent calculations by Mulliken (1948*b*) and Löwdin (1948) reinforce this belief. This argument is concerned, of course, solely with what happens in non-polar molecules, as a result of the geometrical arrangement of atoms around any given atom. It is well recognized that in polar molecules there are other effects which become important, and to which we have not referred here. The most significant of these is the fact that as a result of the charge displacements in such molecules, the Coulomb terms α depend on the final charge distribution, and may not be taken as constants in the way that we have done.

There is another reason for preferring our new correlation with quantities of physical significance. This may be seen most conveniently if we write down the values of H_μ^ν for the two particular molecules butadiene and benzene. The numerical values shown below are calculated on the assumptions that only nearest-neighbour interactions need be considered, and that all $H_{\mu\mu} = \alpha$, $H_{\mu\nu} = \beta$, $S_{\mu\mu} = 1$, $S_{\mu\nu} = 0.25$ (Mulliken, Rieke & Brown 1941). In these expressions γ stands for the combination

$$\gamma = \beta - \alpha S. \quad (39)$$

butadiene:

$$S_{\mu\nu} = \begin{pmatrix} 1 & S & . & . \\ S & 1 & S & . \\ . & S & 1 & S \\ . & . & S & 1 \end{pmatrix}, \quad H_{\mu\nu} = \begin{pmatrix} \alpha & \beta & . & . \\ \beta & \alpha & \beta & . \\ . & \beta & \alpha & \beta \\ . & . & \beta & \alpha \end{pmatrix}$$

$$H_\mu^\nu = \begin{pmatrix} \alpha - 0.287\gamma & 1.148\gamma & -0.305\gamma & 0.076\gamma \\ 1.148\gamma & \alpha - 0.594\gamma & 1.224\gamma & -0.305\gamma \\ -0.305\gamma & 1.224\gamma & \alpha - 0.594\gamma & 1.148\gamma \\ 0.076\gamma & -0.305\gamma & 1.148\gamma & \alpha - 0.287\gamma \end{pmatrix}.$$

benzene:

$$S_{\mu\nu} = \begin{pmatrix} 1 & S & . & . & . & S \\ S & 1 & S & . & . & . \\ . & S & 1 & S & . & . \\ . & . & S & 1 & S & . \\ . & . & . & S & 1 & S \\ S & . & . & . & S & 1 \end{pmatrix},$$

$$H_{\mu\nu} = \begin{pmatrix} \alpha & \beta & . & . & . & \beta \\ \beta & \alpha & \beta & . & . & . \\ . & \beta & \alpha & \beta & . & . \\ . & . & \beta & \alpha & \beta & . \\ . & . & . & \beta & \alpha & \beta \\ \beta & . & . & . & \beta & \alpha \end{pmatrix},$$

$$H_\mu^\nu = (\alpha - 0.622\gamma, 1.241\gamma, -0.356\gamma, 0.178\gamma, -0.356\gamma, 1.241\gamma).$$

Only the top rows of the matrices for benzene need to be given; subsequent ones are obtained by simple permutation.

One interesting thing about these H_μ^ν is that the off-diagonal elements are all simple multiples of $\gamma = \beta - \alpha S$. This is a point in their favour, for Mulliken (1948a) has recently stressed the importance of this expression, originally introduced by Mulliken *et al.* (1941). Another pleasing thing about these figures is the variation shown in the leading diagonal elements, which are all of the form $\alpha - k\gamma$, where k has the values 0.287, 0.594 for C_1 and C_2 in butadiene, and 0.622 for all the carbons in benzene. The fact that the diagonal element is numerically larger at the end atoms than at the central ones in butadiene is in good accord with the calculations of Mulliken (1948b), who showed, from a comparison of graphite, benzene and ethylene, where each carbon atom has 3, 2 or 1 other carbon neighbours, that the more surrounded an atom was with similar neighbours, the smaller would be the numerical value of its diagonal element in the secular determinant.

BOND AND CHARGE AFFINITIES

We have just demonstrated the relevance of the components H_μ^ν of the mixed tensor. It is desirable that we should have some name by which we may refer to them, and distinguish them from the $H_{\mu\nu}$. We suggest that H_μ^μ should be called the *charge affinity* of atom μ , and H_μ^ν the *bond affinity* of μ towards ν . Since $H_\mu^\nu \neq H_\nu^\mu$, we have to be careful in the use of these bond affinities. Our convention is that the lower suffix indicates the atom from which the bond is regarded as starting, and that the upper suffix indicates the direction towards which the bond $\mu - \nu$ is directed. Our reasons for the choice of these names will appear shortly.

Just as the one resonance integral $H_{\mu\nu}$ has had to be replaced by two bond affinities H_μ^ν and H_ν^μ , so we may split up the bond order $p_{\mu\nu}$ into two parts p_μ^ν and p_ν^μ associated with the two ends μ and ν of the bond. The definitions of these quantities (see (33) and (34)) are

$$p_\mu^\nu = \frac{\partial \mathcal{E}}{\partial H_\mu^\nu}, \quad p_\nu^\mu = \frac{\partial \mathcal{E}}{\partial H_\nu^\mu}, \quad q_\mu = \frac{\partial \mathcal{E}}{\partial H_\mu^\mu}. \quad (40)$$

The bond order $p_{\mu\nu}$ is the average of the two parts which we have just introduced, so that

$$p_{\mu\nu} = \frac{1}{2}(p_\mu^\nu + p_\nu^\mu). \quad (41)$$

With these definitions, the total energy of the molecule may be written

$$\mathcal{E} = \sum_{\mu=1}^n q_\mu H_\mu^\mu + \sum_{\mu, \nu=1}^n p_\mu^\nu H_\mu^\nu. \quad (42)$$

This last equation could have been arrived at in a different way. For, by (31) and (11), if the m.o. is normalized:

$$E_j = \sum_{\mu, \nu} x_{j\mu} y_{j\mu} H_\mu^\mu.$$

Thus

$$\begin{aligned} \mathcal{E} &= \sum_{\mu, \nu} n_j x_{j\mu} y_{j\mu} H_\mu^\mu \\ &= \sum_{\mu} q_\mu H_\mu^\mu + \sum_{\mu, \nu} p_\mu^\nu H_\mu^\nu, \quad \text{as required.} \end{aligned}$$

Equation (42) provides the justification for our choice of the names charge- and bond-affinity. Thus, if we have a single charge on atom μ this contributes H_μ^μ to the

energy. This is obviously closely related to what is usually called the electron affinity of atom μ . We have therefore called it the charge affinity of atom μ . Since H_μ^μ —unlike the conventional electron affinity—depends on the other atoms of the molecule as well as on atom μ , it would be more strictly accurate to refer to it as the charge affinity of μ in its position in the given molecule. Exactly similar considerations justify calling H_μ^ν the bond affinity of μ towards ν .

It will be noticed from the table of H_μ^ν for butadiene and benzene given earlier, that bond affinities exist even between atoms so far apart that we can neglect their overlap $S_{\mu\nu}$. Examples are the *para* positions in benzene and the 1, 4 positions in butadiene. It is certainly in accord with chemical intuition that there should be an interaction of this kind and that it should contribute to the energy of the molecule. Nor is it surprising that some of these H_μ^ν are positive and others negative; and that in a conjugated polyene chain their magnitudes decrease as μ and ν are more distant from one another.

Since, from (28), H_μ^ν involves $S^{\mu\nu}$, and (10) shows that $S^{\mu\nu}$ is the inverse matrix of $S_{\mu\nu}$, there is in general no simple expression for H_μ^ν . But in the case of a polyene chain of $2n$ similar atoms numbered consecutively from 1 to $2n$, it may be shown that, neglecting all overlap except between nearest neighbours,

$$H_\mu^\nu = (-1)^{\mu+\nu+1} 4\gamma \cosh^2 \theta \sinh \mu \theta \sinh (2n - \nu + 1) \theta / \sinh \theta \sinh (2n + 1) \theta, \quad (43)$$

where $\cosh \theta = 2$ and it is supposed that $\mu < \nu$. The case of $\mu > \nu$ is obtained at once from the fact that in this problem H_μ^ν is symmetrical in μ and ν . The formula (43) illustrates the alternating character of the H_μ^ν . It also shows that H_μ^ν decreases if ν increases or if μ decreases, corresponding to an increased separation of μ and ν along the chain. It may be added that (43) applies equally, with n instead of $2n$, for even or odd values of n .

POLARIZABILITIES

The bond and charge affinities just discussed enable us to generalize the various self- and mutual-polarizabilities which were introduced in I, and which measure the changes produced in charges or bond orders when any of these affinities is altered.

(i) The atom-atom polarizability gives the change in charge density q_μ induced by a change in the charge affinity H_ν^ν . Thus (cf. I, equation (46))

$$\pi_{\mu,\nu} = \frac{\partial q_\mu}{\partial H_\nu^\nu} = \frac{\partial^2 \mathcal{E}}{\partial H_\mu^\mu \partial H_\nu^\nu} = \pi_{\nu,\mu}. \quad (44)$$

(ii) The bond-atom polarizability gives the change in bond order $p_{\rho\sigma}$ induced by a change in charge affinity H_μ^μ . Thus

$$\pi_{\rho\sigma,\mu} = \frac{\partial p_{\rho\sigma}}{\partial H_\mu^\mu}. \quad (45)$$

(iii) The atom-bond polarizability is

$$\pi_{\mu,\rho\sigma} = \frac{\partial q_\mu}{\partial H_\rho^\sigma} + \frac{\partial q_\mu}{\partial H_\sigma^\rho} = \frac{\partial^2 \mathcal{E}}{\partial H_\rho^\sigma \partial H_\mu^\mu} + \frac{\partial^2 \mathcal{E}}{\partial H_\sigma^\rho \partial H_\mu^\mu} = 2\pi_{\rho\sigma,\mu}. \quad (46)$$

(iv) Finally the bond-bond polarizability is

$$\pi_{\mu\nu,\rho\sigma} = \frac{\partial p_{\mu\nu}}{\partial H_\rho^\sigma} + \frac{\partial p_{\mu\nu}}{\partial H_\sigma^\rho} = \pi_{\rho\sigma,\mu\nu}. \quad (47)$$

The formal resemblance between (44) to (47) and the corresponding equations of I shows that integral expressions completely analogous to I, (52) to (54), can be written down for these quantities, in the manner of (36) and (37). We do not need to reproduce these here.

There is one point concerning equations (44) to (47) which needs further discussion. We have assumed, in making the differentiations, that all the H_μ^ν are independent variables. This assumption was also involved in (32) to (34) and (35) to (38). At first sight this seems an incorrect procedure because the H_μ^ν are derived (see (28)) from the $H_{\mu\nu}$, and there are only $n(n+1)/2$ independent quantities of this kind, on account of the symmetry relation $H_{\mu\nu} = H_{\nu\mu}$. Yet we have treated all n^2 of the H_μ^ν as independent. The justification of our procedure lies in the fact that a complete description of the molecule requires both $H_{\mu\nu}$ and $S_{\mu\nu}$. Now there are $n(n-1)/2$ coefficients $S_{\mu\nu}$ which may be changed by perturbations in the molecule. This is because $S_{\mu\mu} = 1$, and $S_{\mu\nu} = S_{\nu\mu}$. Thus the total number of independent parameters is

$$n(n+1)/2 + n(n-1)/2 = n^2,$$

and we may equally well choose these to be either (a) $H_{\mu\nu}$ and $S_{\mu\nu}$, or (b) H_μ^ν . We have chosen to regard the H_μ^ν as of physical significance rather than the $H_{\mu\nu}$, so the alternative (b) is naturally the one to adopt. We can thus treat all the H_μ^ν as independent parameters, and must realize that the small changes, or perturbations, involved in (40) to (42) could, if we wished, be regarded as small changes both in the $H_{\mu\nu}$ and in the $S_{\mu\nu}$.

There are, of course, alternative expressions for the polarizabilities (44) to (47) in terms of the coefficients $x_{j\mu}$ and $y_{j\mu}$ of the allowed molecular orbitals. These expressions, which are often more convenient to use in a particular example than are the integral representations, may be found by first-order perturbation theory in a manner somewhat similar to that of I, p. 53. On account of the non-orthogonality of the ϕ_μ , the analysis is a little more complicated.

We start with the original secular equations (30) involving the H_ν^μ ; and we suppose that each H_ν^μ receives an increment $H_\nu^{\prime\mu}$, the corresponding increments in $x_{j\nu}$ and $y_{k\mu}$ being $x_{j\nu}'$ and $y_{k\mu}'$. All the 'dashed' quantities are small. Then by straightforward, though somewhat tedious, algebra in which we consider merely terms of the first order of small quantities, it may be shown that

$$x_{j\mu}' = \sum_{k, \nu, \alpha} \frac{x_{j\nu} y_{k\alpha} x_{k\mu} H_\nu^{\prime\alpha}}{E_j - E_k}, \quad (48)$$

and

$$y_{j\mu}' = \sum_{k, \nu, \alpha} \frac{x_{k\nu} y_{j\alpha} y_{k\mu} H_\nu^{\prime\alpha}}{E_j - E_k}. \quad (49)$$

In the course of the proof, which we shall not reproduce in full, we require to use the facts that

$$\sum_\nu x_{j\nu} y_{k\nu} = \sum_\nu x_{k\nu} y_{j\nu}, \quad (50)$$

$$\sum_{\nu, \mu} x_{j\nu} H_\nu^\mu y_{k\mu} = \sum_{\nu, \mu} x_{k\nu} H_\nu^\mu y_{j\mu}, \quad (51)$$

$$\sum_k x_{k\mu} y_{k\nu} = \delta_\mu^\nu. \quad (52)$$

The first two of these may be proved from the transformation properties (9) and (28) for the $x_{j\mu}$ and $y_{j\mu}$. The third is a consequence of geometrical relationships in the space defined by the x_μ , such as those involved in proving equation (12); it may, however, be proved without difficulty by proper use of the orthogonality and normalization equations (11) and (12).

Once we have calculated the changes in the quantities $x_{j\mu}$, $y_{j\mu}$ due to the perturbation, it is easy to determine the changes in any q_μ or $p_{\mu\nu}$, using the definitions of these quantities given in (13) and (14). The method is exactly analogous to I, pp. 53–55, and the results are:

$$\pi_{\mu,\nu} = 2 \sum_{j=1}^m \sum_{k=1}^n \frac{x_{j\nu} x_{k\mu} y_{j\mu} y_{k\nu} + x_{j\mu} x_{k\nu} y_{j\nu} y_{k\mu}}{E_j - E_k} = \pi_{\nu,\mu}, \quad (53)$$

$$\pi_{\mu,\rho\sigma} = 2 \sum_{j=1}^m \sum_{k=1}^n \frac{x_{k\mu} y_{j\mu} (x_{j\rho} y_{k\sigma} + x_{j\sigma} y_{k\rho}) + x_{j\mu} y_{k\mu} (x_{k\rho} y_{j\sigma} + x_{k\sigma} y_{j\rho})}{E_j - E_k} = 2\pi_{\rho\sigma,\mu}, \quad (54)$$

$$\pi_{\mu\nu,\rho\sigma} = \sum_{j=1}^m \sum_{k=1}^n \frac{(x_{j\mu} y_{k\nu} + x_{j\nu} y_{k\mu}) (x_{k\rho} y_{j\sigma} + x_{k\sigma} y_{j\rho}) + (x_{k\mu} y_{j\nu} + x_{k\nu} y_{j\mu}) (x_{j\rho} y_{k\sigma} + x_{j\sigma} y_{k\rho})}{E_j - E_k} = \pi_{\rho\sigma,\mu\nu}. \quad (55)$$

Since all these formulae have numerators that are symmetrical in the suffixes j and k , and denominators that are antisymmetrical, the k -summation need only be taken over the unoccupied orbitals, so that k runs from $m+1$ to n .

These formulae all reduce to the corresponding values (64) to (67) already found in I, when $S_{\mu\nu}$ is neglected; for then $x_{j\mu} = y_{j\mu} = c_{j\mu}$. Further, when $\mathbf{S}$ and $\mathbf{H}$ commute, so that bond orders and charges are independent of overlap, the numerators all have the same values as in I, and so the only differences arise from the altered denominators. The particular case of a hydrocarbon with nearest-neighbour interaction is sufficiently important to be worth describing a little more fully. According to (25) and (26)

$$\epsilon_j = \alpha + m_j \beta, \quad E_j = \frac{\alpha + m_j \beta}{1 + m_j S} = \lambda_j^2 \epsilon_j,$$

$$\text{so that} \quad \epsilon_j - \epsilon_k = (m_j - m_k) \beta, \quad (56)$$

and, after a little reduction,

$$E_j - E_k = \lambda_j^2 \lambda_k^2 (m_j - m_k) \gamma \quad [\gamma = \beta - \alpha S]. \quad (57)$$

Hence the formulae for such hydrocarbons are merely modified by the insertion of $\lambda_j^2 \lambda_k^2$ in the denominators of each term, the result now being given as a multiple of $1/\gamma$ instead of $1/\beta$ as in the earlier theory.

Tables 1 and 2 show the atom-atom polarizabilities calculated according to (53) for benzene and butadiene, the values of S and H being the same as on p. 204.

TABLE 1. ATOM-ATOM POLARIZABILITIES IN BENZENE

	$\pi_{1,1}$	$\pi_{1,2}$	$\pi_{1,3}$	$\pi_{1,4}$
new value, with overlap, units of $1/\gamma$	0.359	-0.137	0.0081	-0.100
old value, without overlap, units of $1/\beta$	0.398	-0.157	0.0093	-0.102

TABLE 2. ATOM-ATOM POLARIZABILITIES IN BUTADIENE

$\pi_{\mu,\nu}$	new values, including S		old values, without S	
	$\nu=1$	$\nu=2$	$\nu=1$	$\nu=2$
$\mu=1$	0.601	-0.377	0.626	-0.402
$=2$	-0.377	0.377	-0.402	0.402
$=3$	0.042	-0.042	0.045	-0.045
$=4$	-0.266	0.042	-0.268	0.045

It is very interesting to find that, notwithstanding the new definitions of polarizabilities, there is really extremely little difference between the two sets of figures. This is particularly gratifying because it justifies us in making our calculations of polarizabilities using the simpler theory, and believing that qualitative conclusions therefrom will be quite unaltered, and quantitative ones only slightly affected. In view of the more clumsy character of the complete theory, the changes are scarcely great enough to warrant its use, except in special cases where considerable accuracy is required.

REFERENCES

- Bradburn, M., Coulson, C. A. & Rushbrooke, G. S. 1948 *Proc. Roy. Soc. Edinb. A*, **62**, 336.
 Coulson, C. A. & Longuet-Higgins, H. C. 1947 *Proc. Roy. Soc. A*, **191**, 39. (Paper I of this series.)
 Coulson, C. A. & Longuet-Higgins, H. C. 1948 *Proc. Roy. Soc. A*, **195**, 188. (Paper V of this series.)
 Longuet-Higgins, H. C. & Coulson, C. A. 1947 *Trans. Faraday Soc.* **43**, 87.
 Löwdin, P.-Olov. 1948 *Ark. Math. Astr. Fys. A*, **35**, no. 30.
 Mulliken, R. S. 1948a International Conference, Paris, published in *J. Chim. Phys.* **46**, 497.
 Mulliken, R. S. 1948b *Phys. Rev.* **74**, 736.
 Mulliken, R. S., Rieke, C. A. & Brown, W. G. 1941 *J. Amer. Chem. Soc.* **63**, 41.
 Wheland, G. W. 1941 *J. Amer. Chem. Soc.* **63**, 2025.

The specific energy of crystal boundaries in tin

BY K. T. AUST AND B. CHALMERS

Department of Metallurgy, University of Toronto

(Communicated by E. N. da C. Andrade, F.R.S.—Received 21 September 1949)

[Plates 10 and 11]

Specimens of tin, consisting of three crystals of predetermined orientations, were prepared with various differences of orientation between two of the crystals. The equilibrium angles between the three boundaries were measured.

Relative values of specific surface free energy were deduced from the angles, and it was found that the surface energy decreases progressively as the difference of orientation decreases below about 6° . The curve relating surface energy with differences of orientation extrapolates to zero surface energy at zero angle. The result is considered to provide strong support for the 'transitional lattice' theory of the crystal boundary.

1. INTRODUCTION

As is well known, a metal consists of crystals which abut against each other at regions known as the crystal boundaries. In a pure metal, the only difference between a crystal and a crystal boundary is the geometrical distribution of the atoms. The structure of the crystal approaches a condition of geometrical perfection, and it is evident that the structure at the boundary must be less perfect.

Direct evidence as to the disposition of the atom at the boundary is not available, as the volume of metal concerned is too small for examination by X-ray methods. Experimental evidence as to the nature of crystal boundaries must, therefore, be somewhat indirect; it must be based on measurements of properties of the material in question. Theoretical considerations have suggested that the boundary consists of a transitional zone where the positions of the atoms are dictated by the force fields arising from both crystal lattices. It is, therefore, theoretically possible to calculate the position of each atom. The only alternative theory is that the boundary region consists of an amorphous layer in which the atoms are distributed in a random fashion; the positions of the atoms would in this case be statistically determinate, but the position of any one atom could not be predicted.

If an intrinsic property of a boundary were found to depend on the relative orientation of the two crystals, this would constitute incontrovertible evidence that different boundaries differ in structure and cannot therefore be amorphous. The fundamental nature of this criterion was pointed out by Chalmers in 1937, and experiments were made (Chalmers 1940) which showed that the melting-point of crystal boundaries in tin was lower than that of the crystals by an amount which is independent of the difference in orientation of the two crystals.

More recently, the possibility of comparing the surface energies of different crystal boundaries has been discussed by Smith (1948), Chalmers (1949) and Dunn & Lionetti (1949). Smith shows that in alpha brass there is some evidence of a statistical character that the angles at which crystal boundaries meet are not all the same, and

that it may therefore be concluded that all the boundaries do not have equal surface energies.

Chalmers (1949) shows that over a considerable range of orientation differences, all the boundaries in tin have equal energies within quite close limits. The evidence in this work was from the angles between the boundaries formed by three crystals of known orientations. The work of Dunn & Lionetti also dealt with boundaries between three crystals. The material in this case was silicon ferrite (3.5 % silicon). The experimental results given by Dunn & Lionetti are reproduced in figure 1. It will be seen that the only evidence for a decrease in surface energy at small differences in orientation is provided by the point no. 1 and by the assumption that the curve should pass through the zero-zero point. This assumption is based on an implicit acceptance of the transitional lattice theory. The scatter of the points leaves some doubt as to the validity of the line which is shown.

The work described in the present paper was undertaken as an extension of the work of Chalmers (1949) to small differences of orientation.

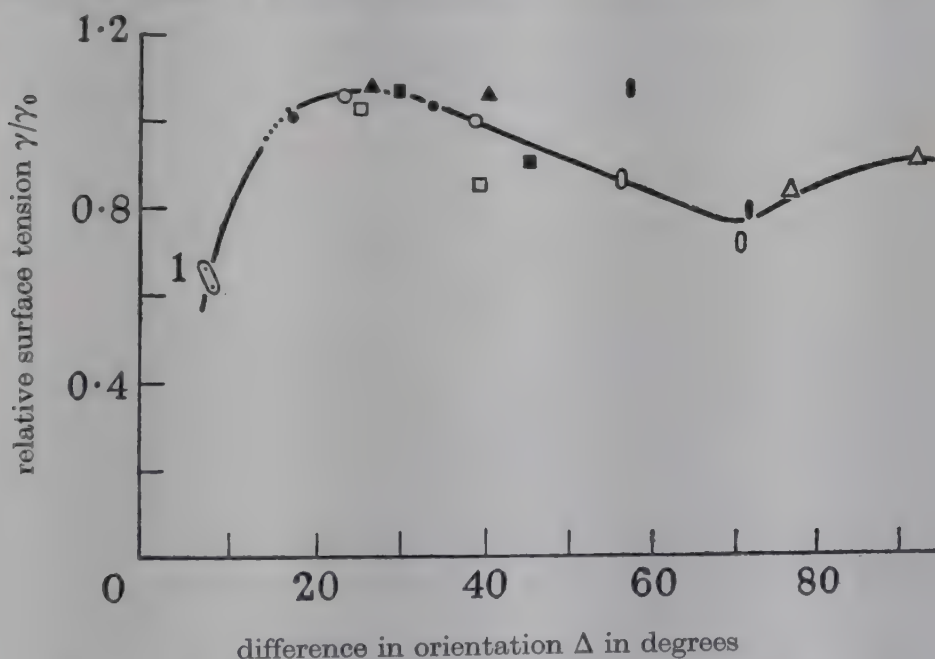


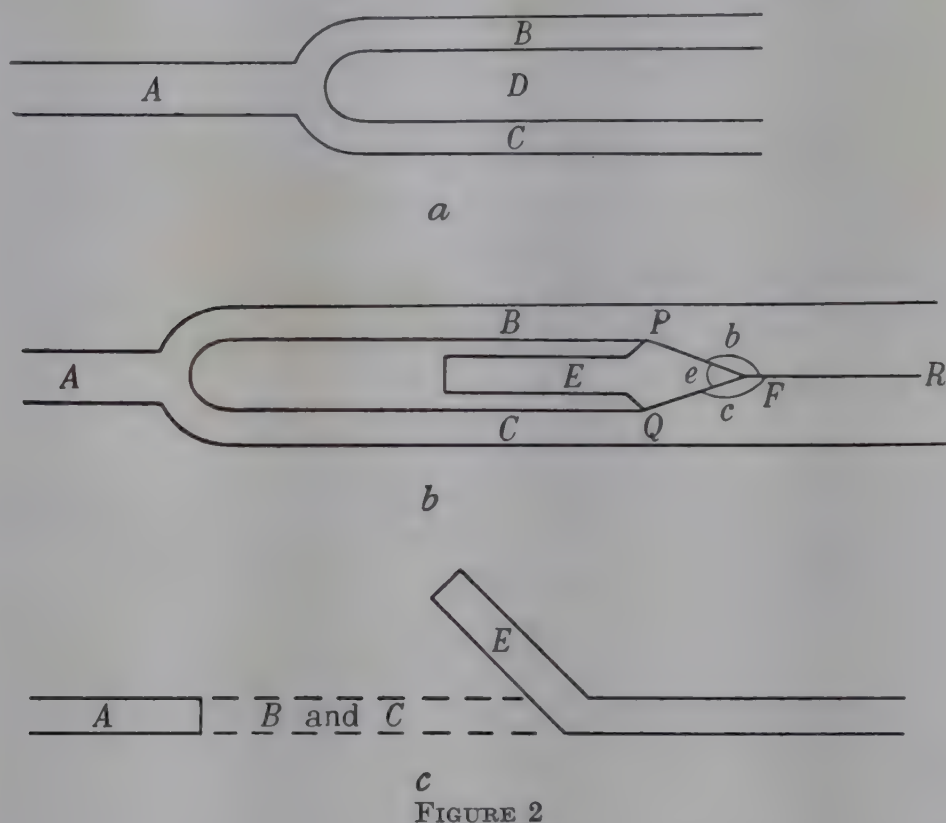
FIGURE 1. Reproduced from Dunn & Lionetti (1949, figure 10).

2. EXPERIMENTAL

The method of preparing the specimen is essentially the same as that described previously. The metal used was tin of 99.987 % purity. The specimens were all grown from 'seed crystals' that were oriented with the C axis perpendicular to the plane of the specimen, and with the a and b axes parallel and perpendicular to the length of the specimen. The first stage of growth consisted of the production of the two rectangular 'legs' B and C from the seed A (figure 2a), using a suitably shaped graphite boat with a removable insert at D .

As grown, the parts B and C have the same orientation; a difference in orientation of the required amount is introduced by rotating one or both of them about their axis through suitable angles. Another seed crystal is then introduced at E . It has the same orientation as A , but it is set at an angle to the plane of the specimen as illustrated in figure 2c. Three seed crystals, B , E and C , are now present side by side,

and the specimen is grown from them. The immediate result is the production of two crystal boundaries (PF and QF), which, by virtue of the orientation of E , converge towards each other. After they meet at F , a third boundary, FR , is formed. Thus, three boundaries meet at the point F ; the difference in orientation across the boundary FR is determined by the rotation of B or C , and the differences across the other two boundaries are virtually constant. A typical specimen is illustrated in figure 3, plate 10.



The differences of orientation were measured roughly by means of the optical reflexion method (Chalmers 1935) and more accurately by means of X-ray back-reflexion Laue photographs, either by superposing exposures from the two crystals (B and C) on the same film, or by comparing two films.

The accuracy of the measurements of the difference of orientation was not better than about $\pm 1^\circ$ because each X-ray photograph showed groups of spots instead of single spots. These are attributed to the 'macro-mosaic' effect described in the previous paper. An example of a back-reflexion photograph is shown in figure 4, plate 10.

The specimens so produced were annealed at $220^\circ\text{C} \pm 2^\circ$ for extended periods. In order to measure the angles at which the boundaries meet at F , it was necessary to polish the specimens electrolytically and to photograph them at magnifications of up to 1000. Typical examples are shown in figures 5 and 6, plate 11.

3. EXPERIMENTAL RESULTS

The initial angles between the boundaries (i.e. during growth of the specimens) were about e , 45° ; b , $157\frac{1}{2}^\circ$; c , $157\frac{1}{2}^\circ$. It is always found that there is some movement of F towards E during the formation of the specimen—that is, the angle e increases

and b and c decrease. During the annealing period, the point F moves towards E and the angles change progressively, approaching equilibrium values.

The rate of movement of F diminishes with time; an example is shown in figure 7. A similar curve for the change of the angle e with annealing time is also shown in the same figure. It is evident that the angles between the boundaries reach equilibrium values before the point F reaches its equilibrium position.

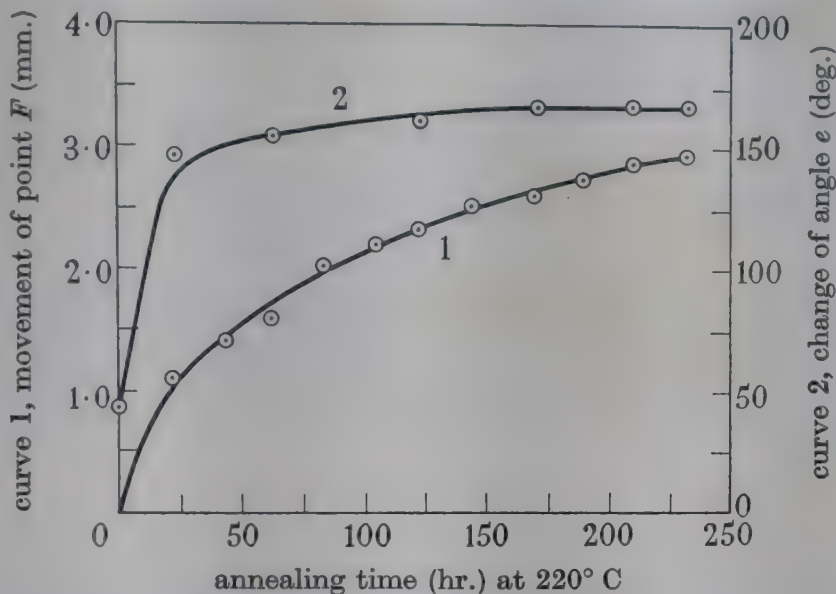


FIGURE 7

Observations were made on a series of specimens in order to determine the equilibrium values of the three angles e , b and c . Annealing was continued until the same angles were recorded after several annealing periods of 20 hr. or more, the total times of annealing amounting to at least 180 hr. The results of these observations are shown in table 1.

TABLE 1

specimen number	total annealing time (hr.) at 220° C	orientation difference (deg.)	angles (deg.)		
			e	c	b
1	290	14	120	119	121
2	225	12	121	121	118
3	181	11	118	119	123
4	332	11	118	120	122
5	342	10	119	121	120
6	181	8	118	122	120
7	225	8	119	118	123
8	191	7	118	124	118
9	202	6	122	116	122
10	288	4½	137	118	105
11	282	4	145	117	98
12	303	3½	147	119	94
13	230	2½	149	106	105
14	231	1	168	93	99

4. DISCUSSION

It will be seen from table 1 that, for most of the specimens, the angles b and c are nearly equal, as would be expected from the symmetry of the orientations of the three crystals. In three cases, however, there is marked divergence between the values of these angles. These are all associated with large values of the angle e , and may be due to the difficulty that is encountered in accurate observation of the boundary FR when it corresponds to a small difference of orientation and, hence, to a low energy. In addition, it may be that low-energy boundaries are influenced to a significant extent by the orientations of the crystals, the energy depending on the orientation of the boundary as well as on the difference of the orientations of the crystals. This possibility has been discussed by Dunn & Lionetti (1949). The uncertainty introduced by this asymmetry is not sufficient to invalidate the main conclusions.

The equilibrium values for the angles were used for calculation of the relative surface energies, the mean surface energy of the boundaries FP and FQ being taken as unity. This is justified by the observation of Chalmers (1949) that the surface energy is constant (for tin) for differences of orientation which are not small; it is not completely in accordance with the results of Dunn, which were, however, obtained with a different material. The results are shown in table 2 and figure 8.

It will be observed that a marked reduction in surface energy takes place when the difference of orientation drops below about 6° , and that the decrease is progressive. A reasonable extrapolation shows zero surface energy for a difference of zero degrees.

It has been assumed in the foregoing discussion that the angles under consideration are at their equilibrium values. This is justified by the fact that in some cases the final angle was reached by an increase of a smaller angle, while in other cases the angle, originally greater, decreases to the final value. The final values showed a satisfactory agreement, and it follows that there is a single minimum in the free-energy angle relationship and that metastable equilibrium states are not reached.

It may, therefore, be concluded that the specific free energy of a crystal boundary in tin depends on the difference of orientation at the boundary when this is less than about 6° and that the energy will be close to zero for zero difference of orientation.

This result may be regarded as confirming the general conclusions of both Smith (1948) and of Dunn & Lionetti (1949) that the energy of a crystal boundary is not independent of the relative orientations of the neighbouring crystals.

TABLE 2

orientation difference (deg.)	relative surface energy	orientation difference (deg.)	relative surface energy
14	1.00	7	1.03
12	0.99	6	0.97
11	1.03	$4\frac{1}{2}$	0.73
11	1.03	4	0.60
10	1.01	$3\frac{1}{2}$	0.57
8	1.03	$2\frac{1}{2}$	0.54
8	1.01	1	0.21

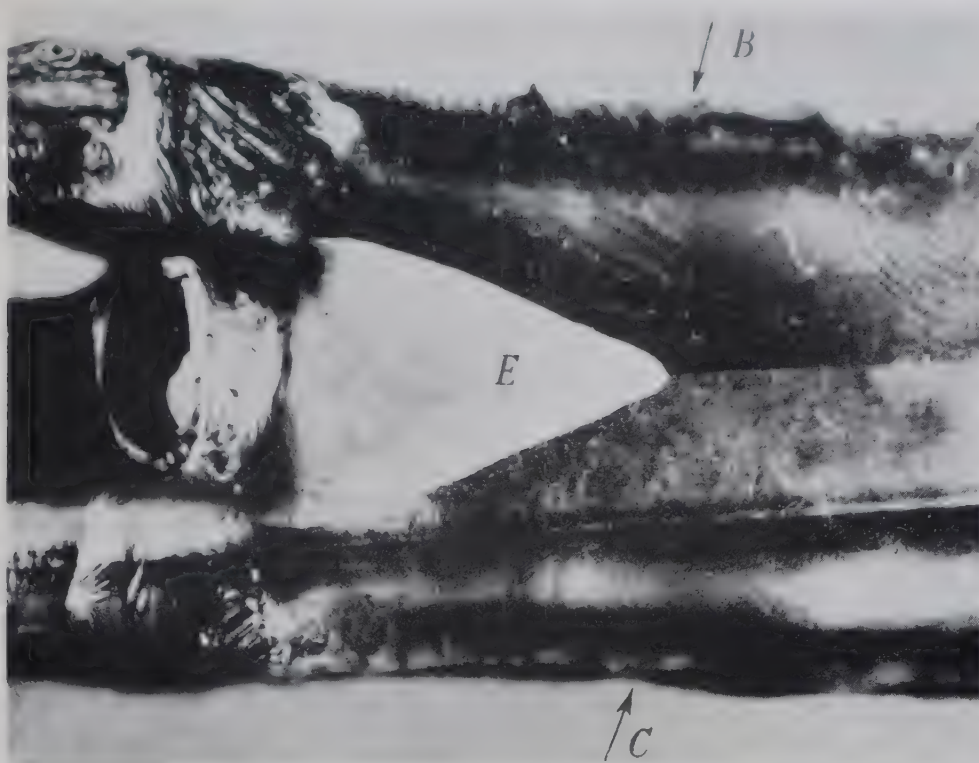


FIGURE 3

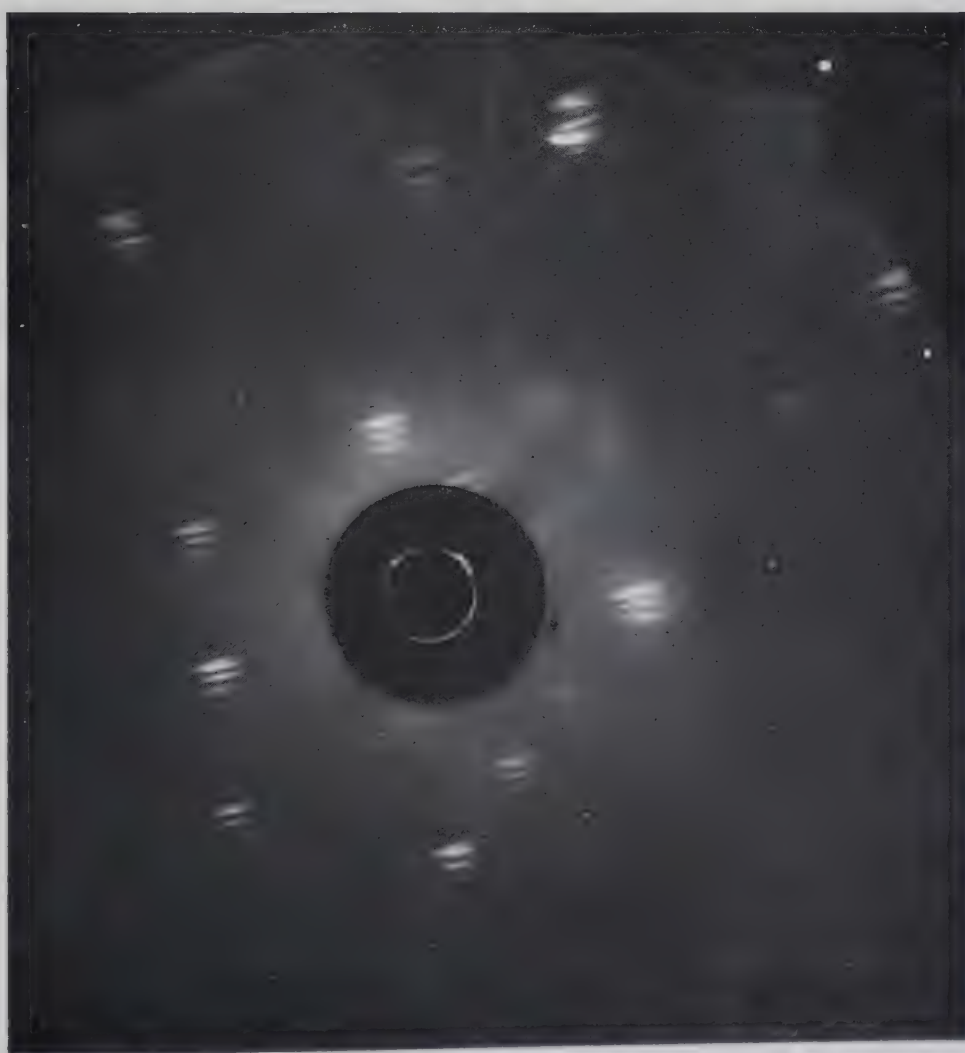


FIGURE 4

(Facing p. 214)



FIGURE 5. Specimen no. 8 after 191 hr. at 220° C; electropolished. Magnification $\times 300$.

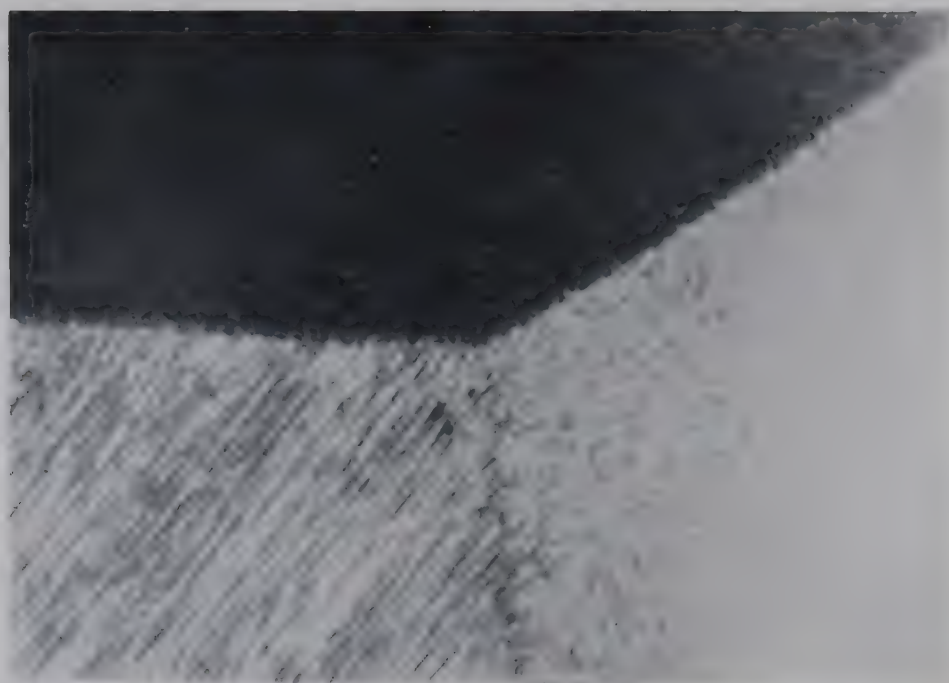


FIGURE 6. Specimen no. 11 after 203 hr. at 220° C; electropolished. Magnification $\times 300$.

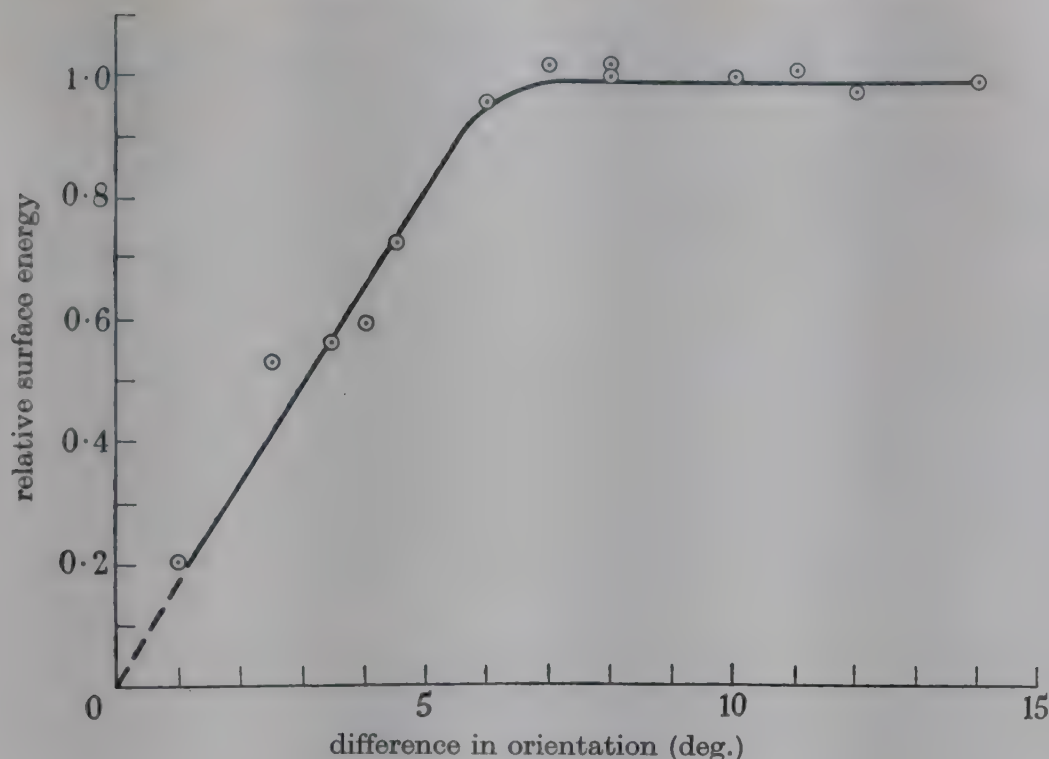


FIGURE 8

It is the authors' opinion that the results described in this paper present the most convincing experimental evidence that has yet been produced that an intrinsic property of a crystal boundary depends on the crystallographic relationship of the crystals between which it is formed. It is, therefore, considered to furnish very strong support for the transitional lattice model of the crystal boundary.

The authors wish to express their indebtedness to the National Research Council, Ottawa, the Research Council of Ontario, Toronto, and the School of Engineering Research of the University of Toronto, for financial assistance.

REFERENCES

- Chalmers, B. 1935 *Proc. Phys. Soc.* **47**, 733.
 Chalmers, B. 1937 *Proc. Roy. Soc. A*, **162**, 120.
 Chalmers, B. 1940 *Proc. Roy. Soc. A*, **175**, 100.
 Chalmers, B. 1949 *Proc. Roy. Soc. A*, **196**, 64.
 Dunn, C. G. & Lionetti, F. 1949 *J. Metals Tech.* **1**, 125.
 Smith, C. S. 1948 *Metals technology*, June.

Cellular atmospheric waves in the ionosphere and troposphere

BY D. F. MARTYN, D.Sc., *Commonwealth Observatory, Canberra*

(Communicated by S. Chapman, F.R.S.—Received 6 October 1949)

Consideration of horizontally travelling disturbances of electron densities in the F_2 region of the ionosphere suggests that they are identical with the supposed vertically travelling disturbances discovered by Wells, Watts & George (1946). They appear to be due to horizontally travelling atmospheric cellular waves of the type first investigated by Lamb. The theory of such waves is developed to include wind with linear vertical shear, special attention being paid to the conditions of bounding by temperature gradients. It is found that such waves appear to be the cause of the microbarometric oscillations long known in the troposphere. The theory of these is worked out in some detail, and an explanation found for Johnson's discovery that these oscillations have periods agreeing closely with Brunt's natural period of oscillation of a *small element* of atmosphere.

In the ionosphere, the earth's magnetic field profoundly affects the observed motions of ionization in the F_2 region, leading to a height gradient in the phase of these oscillatory motions. This gradient makes horizontally travelling disturbances appear to have a vertical component of motion, thus simulating vertically moving electron clouds. The bounding conditions for these cells in the F_2 region appear to necessitate a value for γ , the ratio of specific heats of air, considerably less than 1.4.

1. INTRODUCTION

This investigation began in a search for the explanation of a recently discovered type of travelling disturbance (Munro 1948; Beynon 1948) in the ionosphere. As it developed, it became evident that the principles involved have some features in common with other atmospheric phenomena of the troposphere, notably the microbarometric oscillations, billow clouds, waves in the lee of mountains, and possibly the cells of 'bumpiness' encountered by aircraft. It appears to be increasingly recognized that most of these phenomena involve cellular waves, rotational in type, travelling between two horizontal bounding surfaces. This viewpoint has been reached slowly and with difficulty. It appears that for many years the work of Helmholtz (1889) on waves at a surface of discontinuity of density and wind was regarded as a satisfactory foundation for the origin of billow clouds and microbarometric oscillations. The inadequacy of such waves to account for billow clouds appeared from the work of Terada, Brunt, Walker and others. Recently, the foundation of a satisfactory cellular wave theory of these clouds has been given by Sekera (1948). The theory of waves in the lee of mountains has also been developed very recently by Scorer (1949).

The theory of the microbarometric oscillations remains in an unsatisfactory state. Following Goldie's (1925) treatment of these oscillations as Helmholtz waves, Johnson (1929) showed that their frequency agreed well with the natural frequency of vertical oscillation of a *small element* of atmosphere as deduced by Brunt (1927). While this result suggested a resonance phenomenon, it was generally recognized that no sound theoretical reason appeared why the oscillating air mass should vibrate at the free period of a small element of air, rather than at a free period of the atmosphere as a whole. A recent attempt at the theory of this subject by Pekeris (1948) treats the atmosphere as a whole, with the troposphere of uniform lapse rate, and an isothermal stratosphere. It appears, however, that this theory also is unsatisfactory,

since it leads to velocities of travel approaching that of sound, whereas Goldie (1925) has shown that typical velocities are $\frac{1}{25}$ th of this.

In the present work a theory of these oscillations is developed in some detail. It is found that oscillations of the observed frequencies and velocities can occur as cellular waves between two horizontal surfaces, one of these being the ground, and the other a level at which there occurs a sudden *increase* in either lapse rate or temperature. The possible frequencies normally lie within a very limited range and approximate closely to the Brunt (angular) frequency (σ_B) for the region *above* the discontinuity. It appears that Johnson's result shows, not that the atmosphere oscillates at the frequency σ_B , but rather that it will *not* oscillate at σ_B ; in other words, the level at which the lapse rate and temperature give the atmosphere a local natural frequency σ_B forms a bounding surface for oscillations of this (and higher) frequencies. The lower limit of possible frequencies is provided by the condition that the oscillations must be unbounded between this upper surface and the ground; otherwise no oscillations would be observed.

The application of the same theory to the ionosphere necessitates two bounding surfaces in or near the ionosphere itself. The lower of these *might* occur at about the 80 km. level, where high lapse rates are known to exist (Martyn & Pulley 1936). There is considerable difficulty, however, in locating an upper boundary for the wavelengths and frequencies which are experimentally found. It seems quite unlikely that there is any marked decrease in temperature above the F_2 region (250 km.). Such a postulate would make it difficult to understand the presence of adequate atmosphere at the great heights (> 1000 km.) at which sunlit aurorae occur. The absence of an upper bounding surface would lead to considerable vertical leakage of energy; this is difficult to reconcile with the evidence shown later that these disturbances can travel distances of the order 1000 km. without sensible decay. The only likely alternative appears to be that γ , the ratio of the specific heats of the atmospheric gases, is considerably less than 1.4 in the F_2 region.

2. DERIVATION OF THE WAVE EQUATION

The axis of x is taken horizontal, that of y vertical and positive upwards. In the Eulerian notation, the equations of motion are:

$$\frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} = -\frac{1}{\rho_0} \frac{\partial p}{\partial x}, \quad (1)$$

$$\frac{\partial v}{\partial t} + U \frac{\partial v}{\partial x} = -\frac{1}{\rho_0} \left(\frac{\partial p}{\partial y} + g\rho \right), \quad (2)$$

where u, v are the (small) horizontal and vertical air velocities at x, y at time t ; p and ρ are the pressure and density of air at this point; and p_0, ρ_0 the equilibrium values of p, ρ . The departures from these equilibrium values are given by

$$p' = p - p_0 \quad \text{and} \quad \rho' = \rho - \rho_0.$$

It is assumed that the *variations* of pressure follow the adiabatic relation

$$\frac{Dp}{Dt} = c^2 \frac{D\rho}{Dt}, \quad (3)$$

where

$$c^2 = \gamma p_0 / \rho_0 = \gamma R \theta_0 = \gamma g H,$$

R being the gas constant, θ_0 the equilibrium temperature at y , and H the 'height of the homogeneous atmosphere'. It follows that c is the velocity of sound at y .

The equation of continuity is

$$\frac{D\rho}{Dt} + \rho_0 \chi = 0, \quad (4)$$

where $\chi \left(= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)$ is the hydrodynamical divergence.

Neglecting terms of the second order of smallness, (3) and (4) may be written

$$\frac{\partial p'}{\partial t} + U \frac{\partial p'}{\partial x} + v \frac{\partial p_0}{\partial y} = c^2 \left(\frac{\partial \rho'}{\partial t} + U \frac{\partial \rho'}{\partial x} + v \frac{\partial \rho_0}{\partial y} \right), \quad (3')$$

$$\frac{\partial \rho'}{\partial t} + U \frac{\partial \rho'}{\partial x} + v \frac{\partial \rho_0}{\partial y} + \rho_0 \chi = 0. \quad (4')$$

For a wave travelling in the x direction a factor $e^{i(\sigma t - kx)}$ is assumed in all the dependent variables. Now eliminating p' and ρ' from these equations there results

$$\frac{iG}{kc^2} u = (\Omega^2 - g/H) \chi + g\chi', \quad (5)$$

$$\frac{G}{c^2} v = (k^2 g - \Omega^2/H + k\Omega U') \chi + \Omega^2 \chi', \quad (6)$$

where Ω is $(\sigma - kU)$, G is $k^2 g^2 + (gkU' - \Omega^3)\Omega$, and the prime is used to denote partial differentiation with respect to y .

Differentiating (5) and (6) with respect to y one obtains the differential equation for the divergence, viz.

$$\begin{aligned} \chi'' + \left(\frac{H' - 1}{H} - \frac{kU'}{\Omega} \right) \chi' + \left[-k^2 + \Omega^2/c^2 + \frac{k^2 g}{\Omega^2 H} \left(\frac{\gamma - 1}{\gamma} + H' \right) + \frac{k}{\Omega} \left(\frac{2\gamma - 1}{\gamma H} + \frac{H'}{H} \right) U' \right. \\ \left. + \frac{kU''}{\Omega} - \frac{k^2}{\Omega^2} U'^2 + \frac{k^2}{G} \left\{ g \left(\frac{\Omega}{kH} - \frac{gk}{\Omega} - U' \right) U'' \right. \right. \\ \left. \left. + 4\Omega \left(\frac{\Omega^2}{kH} - kg \right) U' + \left(4\Omega^2 + \frac{g^2 k^2}{\Omega^2} - \frac{g}{H} \right) U'^2 + \frac{gk}{\Omega} U'^3 \right\} \right] \chi = 0. \end{aligned} \quad (7)$$

In terms of χ , p' is given by

$$\frac{iG}{\gamma \Omega} \frac{p'}{p_0} = (\Omega^2 - g/H) \chi + g\chi'. \quad (8)$$

From now on attention is confined to linear gradients of H and U such that

$$H = H_0(1 + \alpha y), \quad U = U_0(1 + \beta y).$$

It is also assumed that $U' < 10^{-2} \text{ sec.}^{-1}$, $\sigma < 10^{-2} \text{ sec.}^{-1}$, and $3 \times 10^{-7} < k < 10^{-4}$, corresponding roughly to wave-lengths between 0.6 and 200 km. Under these restrictions many terms become zero or negligible in equation (7), which reduces to

$$\chi'' + \left(\frac{H' - 1}{H} - \frac{kU'}{\Omega} \right) \chi' + \left\{ -k^2 + \frac{\Omega^2}{c^2} + \frac{k^2 g}{\Omega^2 H} \left(\frac{\gamma - 1}{\gamma} + H' \right) + \frac{k}{\Omega H} \left(\frac{2\gamma - 1}{\gamma} + H' \right) U' \right\} \chi = 0. \quad (7')$$

Substituting

$$y = \int H dz$$

and

$$\chi = \phi e^{z/2} \Omega / (\sigma - kU_{H=0}),$$

there results

$$\frac{\partial^2 \phi}{\partial z^2} + Q^2 \phi = 0, \quad (9)$$

where

$$Q^2 = \left\{ -k^2 H^2 + \frac{k^2 g H}{\Omega^2} \left(\frac{\gamma - 1}{\gamma} + \alpha H_0 + \frac{\beta^2 U_0^2 H}{4g} \right) + \frac{k \beta U_0 H}{2\Omega} \left(\frac{3\gamma - 2}{\gamma} + 3\alpha H_0 \right) + \frac{\Omega^2 H}{\gamma g} - \frac{1}{4} \right\}. \quad (10)$$

Equation (9) may be regarded as expressing the vertical propagation of ϕ with a phase velocity proportional to $1/Q$. Hence boundaries (for ϕ) will occur at levels where $Q^2 = 0$, for at such levels the group velocity of propagation must be zero. It will, of course, be a necessary condition for adequate bounding that Q remain imaginary for a distance which is an appreciable fraction of a wave-length on the far side of such a boundary. The mean rate of transmission of energy upwards is given by the mean value of $\mathcal{R}(p'v^*)$, where $\mathcal{R}$ denotes 'the real part of', and the $*$ is used to denote the complex conjugate. From equations (6) and (8) it follows that

$$\mathcal{R}(p'v^*) = \mathcal{I} \left\{ p \gamma^2 g e^z (\sigma - kU_{H=0}) / G \cdot \phi^* \frac{\partial \phi}{\partial z} \right\}, \quad (11)$$

where $\mathcal{I}$ denotes 'the imaginary part of'.

In the particular case of no winds in the x direction, equation (9) reduces to

$$\frac{\partial^2 \phi}{\partial z^2} + \left\{ -k^2 H^2 + \frac{k^2 g H}{\sigma^2} \left(\frac{\gamma - 1}{\gamma} + \alpha H_0 \right) + \frac{\sigma^2 H}{\gamma g} - \frac{1}{4} \right\} \phi = 0. \quad (9')$$

It is clear from equations (9) and (9'), together with consideration of the magnitudes of the parameters, that, in general, the principal effect of the wind is to alter the frequency of the oscillation *in the moving medium* from σ to $\sigma - kU$. It is obviously possible for this frequency (Ω) to become zero, and so to pass through a range of values wherein it is comparable with the angular frequency of rotation of the earth (ω). It is desirable therefore to examine how this latter frequency is likely to enter the equations. For this purpose the differential equation for χ is developed below for the two cases of disturbances travelling east and north. For the first case, the dynamical equations (for $U_0 = 0$) are

$$\rho_0 \frac{\partial u}{\partial t} + 2\omega(w \cos \Phi - v \sin \Phi) \rho_0 = -\frac{\partial p'}{\partial x}, \quad (12)$$

$$\rho_0 \frac{\partial v}{\partial t} + 2\omega \rho_0 \sin \Phi u = 0, \quad (13)$$

$$\rho_0 \frac{\partial w}{\partial t} - 2\omega \rho_0 \cos \Phi u = -g\rho' - \frac{\partial p'}{\partial z}, \quad (14)$$

where u, v are the air velocities east and north respectively, w is the velocity upwards, and Φ the geocentric latitude. The equation of continuity gives

$$\frac{\partial \rho'}{\partial t} + w \frac{\partial \rho_0}{\partial z} + \rho_0 \chi = 0, \quad (15)$$

while the equation of state (3) when combined with (15) gives

$$\frac{\partial p'}{\partial t} - g\rho_0 w = -\gamma p_0 \chi. \quad (16)$$

Elimination of p' and ρ' from these equations gives

$$\sigma_1 c^2 \frac{\partial^2 \chi}{\partial z^2} + \sigma_1^2 \left(\frac{dc^2}{dz} - g\gamma \right) \frac{\partial \chi}{\partial z} + \left\{ \sigma^4 - 4\sigma^7 \omega^2 + k^2 g^2 (\gamma - 1) - k^2 \sigma^7 c^2 + 2kg\sigma\omega \cos \Phi (2 - \gamma) \right. \\ \left. + k(kg - 2\sigma\omega \cos \Phi) \frac{dc^2}{dz} \right\} \chi = 0, \quad (17)$$

where

$$\sigma_1^2 = \sigma^2 - 4\omega^2 \sin^2 \Phi.$$

Proceeding in the same way the differential equation for a north-travelling disturbance is found to be

$$\sigma_1^2 c^2 \frac{\partial^2 \chi}{\partial z^2} + \left\{ \sigma_1^2 \left(\frac{dc^2}{dz} - g\gamma \right) + ikc^2 (4\omega^2 \sin^2 \Phi) \right\} \frac{\partial \chi}{\partial z} \\ + \left\{ \sigma^4 - 4\sigma^7 \omega^2 + k^2 g^2 (\gamma - 1) - k^2 \sigma^2 c^2 + 4k^2 c^2 \omega^2 \cos^2 \Phi \right. \\ \left. - 2i\gamma kg\omega^2 \sin^2 \Phi + k(kg + 2i\omega^2 \sin^2 \Phi) \frac{dc^2}{dz} \right\} \chi = 0. \quad (18)$$

It is clear that the main effect of the earth's rotation, for the ranges of k and σ under examination, is to alter the square of the effective frequency of oscillation from σ^2 to $(\sigma^2 - 4\omega^2 \sin^2 \Phi)$.

3. DISCUSSION OF BOUNDARY CONDITIONS

We begin with the study of the case of no wind along the direction of propagation (equation (9')). Excluding for the moment the special case of $\sigma = 0$, it is clear that a horizontal bounding surface will be given by the condition $Q_{U=0}^2 = 0$. At this level the phase velocity for vertical propagation is ∞ and the group velocity therefore zero. Beyond this level, for a linear gradient of temperature, the velocity of propagation remains imaginary. For given values of temperature and temperature gradient, the relation between σ and k for bounding may be studied by solving the quadratic equation in σ^2 ,

$$\sigma^4 - \gamma g(k^2 H^2 + \frac{1}{4})/H \sigma^2 + \gamma g^2 k^2 (1 - 1/\gamma + \alpha H_0) = 0. \quad (19)$$

Solutions of this equation for a representative range of values of k , for $H = 8$ km., and with αH_0 corresponding to two-thirds the dry adiabatic lapse rate, are plotted in figure 1. The resulting curves may therefore be considered as representative of average tropospheric conditions. Two curves result, the upper cutting the σ axis at $c/2H$, and approaching the straight line $\sigma/k = c$ asymptotically. For values of σ , k having representative points in the shaded area between this curve and the σ axis it may easily be verified that $Q^2 > 0$ and χ is unbounded. The lower curve starts from the origin and approaches the horizontal line $\sigma = \sqrt{\left\{ \frac{g}{H} \left(\frac{\gamma - 1}{\gamma} + \alpha H_0 \right) \right\}}$ asymptotically.

For small values of k it approximates to the straight line $\sigma/k = 2 \sqrt{\left\{ gH \left(\frac{\gamma - 1}{\gamma} + \alpha H_0 \right) \right\}}$.

In the shaded region between this curve and the k axis χ is unbounded. For lapse rates of temperature greater than $\frac{2}{3}\Gamma$ the lower curve collapses, eventually coinciding with the k axis when the gradient has the full adiabatic lapse rate. For an isothermal atmosphere $\alpha = 0$ and the lower curve rises to approach $\sigma = \sqrt{\left(\frac{g}{H} \frac{\gamma-1}{\gamma}\right)}$ asymptotically. This frequency now closely approaches the lowest frequency $\sigma = c/2H = \sqrt{\left(\frac{g}{H} \frac{\gamma}{4}\right)}$ of the upper curve.

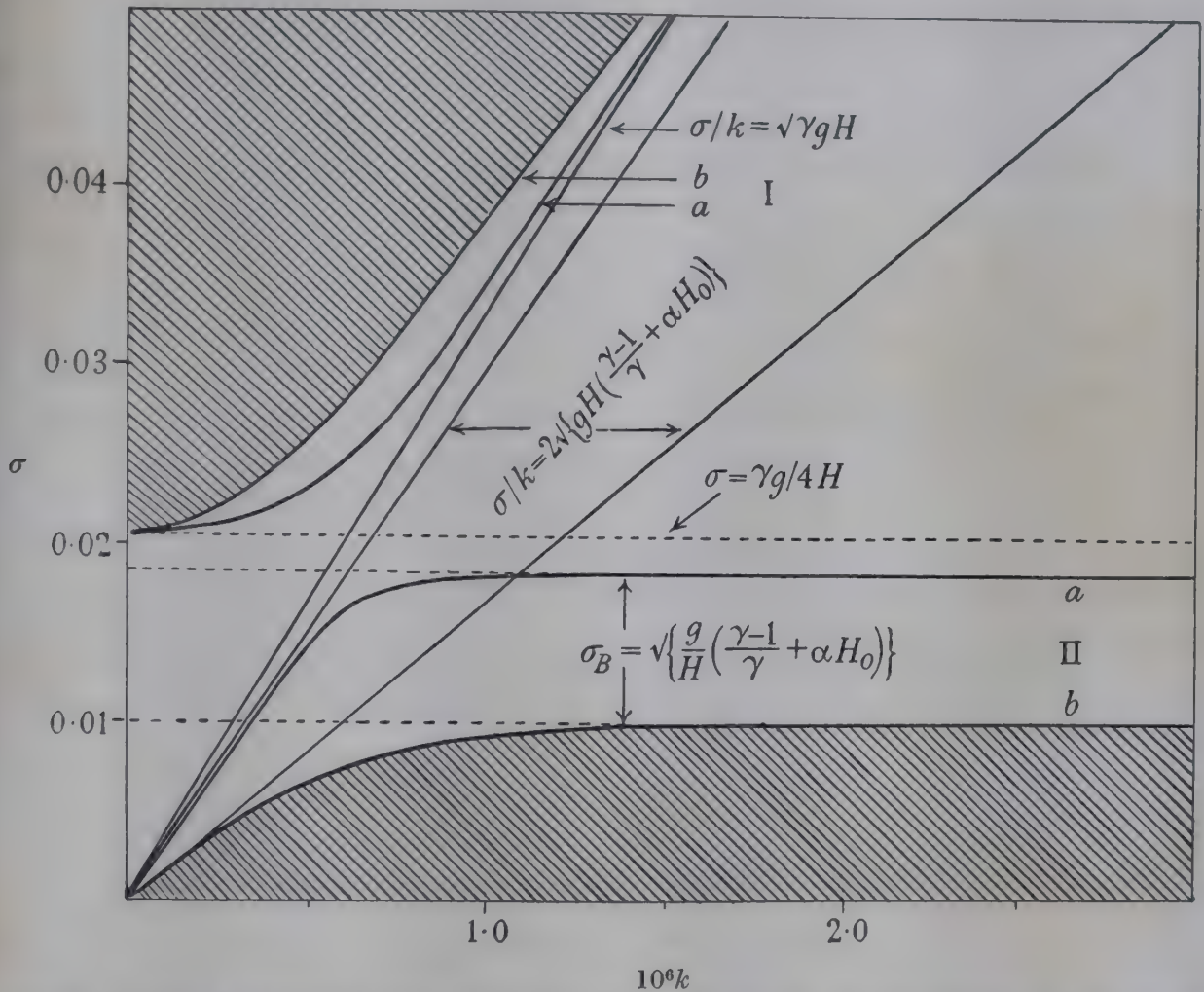


FIGURE 1. The σ, k relations for vertical bounding of cellular atmospheric waves: (a) isothermal atmosphere, (b) lapse rate $\frac{2}{3}$ adiabatic. The hatched areas represent unbounded waves for case (b).

It is clear that the upper part of the upper curve gives the bounding condition for sound waves. The deviation from linearity at low frequencies is obviously associated with the increased phase velocity of vertical propagation of sound occurring when the wave-length becomes comparable with $4\pi H$. The lowest frequency on this curve corresponds to a period of about 5 min. near the ground (Lamb 1908).

The lower curve is obviously that which gives the bounding conditions for the range of phenomena considered in this paper. The horizontal phase velocities of the bounded waves are always less than those of sound waves. In particular, for moderately short wave-lengths (less than some tens of km.) the bounding condition

$$\sigma^2 = \frac{g}{H} \left(\frac{\gamma-1}{\gamma} + \alpha H_0 \right) = \frac{g}{\theta} \left(\Gamma + \frac{d\theta}{dy} \right) = \sigma_B^2$$

is identical with that found experimentally by Johnson (1929) for the frequency of the microbarometric oscillations.

It will be observed that under normal conditions the troposphere acts as a barrier to the vertical propagation of all wave-lengths with periods lying between 5 and 10 min. When the lapse rate is greater than $\frac{2}{3}\Gamma$ this stopping band widens, eventually covering periods from 5 min. to ∞ when $d\theta/dy = -\Gamma$.

It is clear from equations (17) and (18) that consideration of the earth's rotation will introduce only minor modifications to the curves of figure 1. The lower curve will not pass through the origin but will cut the σ axis at $\sigma = 2\omega \sin \Phi$ corresponding to periods of the order of 1 day. Oscillations of period longer than this will be bounded.

In general, it is clear that the most important factor in providing a boundary is the lapse rate. When this approaches Γ all wave-lengths are bounded for periods longer than 5 min. On the other hand, when α is positive, oscillations of all periods less than about a day can propagate vertically in certain wave-length ranges. A secondary factor in providing a boundary is the absolute temperature, or, more generally, H . Large values of H or θ increase the range of bounded frequencies at moderately short wave-lengths (some tens of km.). This effect can be of little importance in the troposphere, but may have significance in the ionosphere, where the temperatures probably range up to one or two thousand degrees Kelvin (Martyn & Pulley 1936).

Turning now to cases where $U_0 \neq 0$ it is seen from equation (10) that σ must be replaced by Ω . Also, several additional terms appear in the expression for Q^2 . For present purposes the last two terms in Q^2 may be neglected. The last term ($-\frac{1}{4}$) is only of importance for long waves, while the second last ($\Omega^2 H/\gamma g$) is important only for sound waves. The third last term, in (10), will also generally be negligible so long as kH is substantially greater than unity. Thus Q^2 simplifies to

$$Q^2 = -k^2 H^2 + \frac{k^2 g H}{\Omega^2} \left(\frac{\gamma - 1}{\gamma} + \alpha H_0 + \frac{\beta^2 U_0^2 H}{4g} \right), \quad (20)$$

so that the condition for bounding becomes

$$\Omega^2 = \frac{g}{H} \left(\frac{\gamma - 1}{\gamma} + \alpha H_0 + \frac{\beta^2 U_0^2 H}{4g} \right). \quad (21)$$

Thus even in the limiting case of adiabatic lapse rate ($\frac{\gamma - 1}{\gamma} + \alpha H_0 = 0$), Ω^2 remains finite and positive. There remains then a range of low frequencies *in the medium* which are unbounded unless the lapse rate be super-adiabatic.

The gradient of U itself provides another bounding condition which may operate both in the troposphere and ionosphere. For values of U ranging from 10^3 to 10^4 cm./sec. kU becomes comparable with σ . The former velocities are commonly encountered in the troposphere, while the evidence of tidal movements (Martyn 1948) and meteor trails (Olivier 1947) shows that the latter velocities are of common occurrence in the upper atmospheric winds. Equation (20) shows that there are two ways in which these winds can produce bounding. If $U < 0$ then the frequency Ω in the medium is raised, and for sufficiently large $|U|$ Q^2 can become negative for quite moderate lapse rates. It seems unlikely, however, that this condition will have much practical importance, since it implies a disturbance travelling against the wind.

Since in most cases the disturbance is likely to be produced by wind shear it will most often travel with the wind. Certainly this is the case for billow clouds.

The second manner in which the wind can produce bounding occurs when the disturbance travels in the same direction as the wind. It is clearly possible in this case for Ω to become zero at some level. The disturbance then travels with the wind velocity and is a state of steady motion with reference to axes moving with this velocity. This condition is of importance in the theory of billow clouds, and the bounding criteria have been fully treated by Sekera (1948). For reasons which will appear below, these cannot be responsible for the bounding of microbarometric oscillations, but appear likely to be important for travelling ionospheric disturbances.

Another condition for bounding is revealed by equation (11). When Ω becomes zero at the level where H is zero then there is no mean flow of energy vertically at any level.

4. SOLUTION OF THE DIFFERENTIAL EQUATION

Equation (9') may be written in the form

$$\frac{\partial^2 \phi}{\partial z^2} + \left(-a e^{2\alpha H_0 z} + b e^{\alpha H_0 z} - \frac{1}{4} \right) \phi = 0, \quad (22)$$

where

$$a = k^2 H_0^2,$$

$$b = \left\{ \frac{gk^2}{\sigma^2} \left(\frac{\gamma-1}{\gamma} + \alpha H_0 \right) + \frac{\sigma^2}{\gamma g} \right\} H_0.$$

For $|z| < 1$, i.e. for ranges of y less than H , $|\alpha H_0 z| \ll 1$ and equation (22) may be written

$$\frac{\partial^2 \phi}{\partial z^2} + (m + nz) \phi = 0, \quad (23)$$

with $m = b - a - \frac{1}{4}$, $n = (b - 2a) \alpha H_0$. Substituting $n = m + nz$, equation (23) becomes

$$\frac{\partial^2 \phi}{\partial \eta^2} + \frac{\eta}{n^2} \phi = 0. \quad (24)$$

This is a standard form of Riccati equation, the solution being expressible in Bessel functions of order $\pm \frac{1}{3}$, viz.

$$\phi = \eta^{\frac{1}{3}} \left\{ A J_{\frac{1}{3}} \left(\frac{2}{3n} \eta^{\frac{3}{2}} \right) + B J_{-\frac{1}{3}} \left(\frac{2}{3n} \eta^{\frac{3}{2}} \right) \right\}, \quad (25)$$

with

$$\frac{\partial \phi}{\partial z} = \frac{\eta}{n} \left\{ A J_{-\frac{1}{3}} \left(\frac{2}{3n} \eta^{\frac{3}{2}} \right) - B J_{\frac{1}{3}} \left(\frac{2}{3n} \eta^{\frac{3}{2}} \right) \right\}. \quad (26)$$

If one boundary is a fixed surface (e.g. the surface of the earth) then the condition to be satisfied at this level is $v = 0$, or

$$(\sigma^2 - 2gk^2 H) \phi - 2\sigma^2 \frac{\partial \phi}{\partial z} = 0. \quad (27)$$

The other types of boundary will occur in the atmosphere at discontinuities of temperature gradient, where Q^2 suddenly becomes negative, or at the level $Q^2 = 0$ when the gradient is uniform. In both cases, provided Q^2 remains negative for an

appreciable fraction of a wave-length, it will be adequate to put $\phi = 0$ at these boundaries. A full wave treatment would lead to the appearance of small oscillating terms in the solutions, at the expense of considerable mathematical complications. By equations (3) and (4) it is seen that this condition ($\phi = 0$) corresponds to $Dp/Dt = 0$, which is the condition for a free surface, the moving particles experiencing no change in pressure in the course of their wave motion on the boundary. Once the boundaries are specified, the arbitrary constant can be eliminated and the relation between σ and k explicitly determined. Tables of Bessel functions of order $\pm \frac{1}{3}$ are given very fully by Watson (1944), while Dinnik (1911) has tabulated $J_{\pm \frac{1}{3}}$ for arguments from 0 to 8.0 in steps of 0.2.

When the oscillation extends vertically through a large part of the atmosphere $|z| \gg 1$, and it is not possible to use the above approximate solution. For such cases it is probably most convenient to follow a method due to Langer (1937). Langer's solution, valid when Q^2 has a simple zero at $z = 0$, and with Q^2 positive for $z > 0$ is

$$\phi \sim \left\{ \frac{8\pi\zeta}{3Q} \right\}^{\frac{1}{2}} \{ \cos(\tfrac{1}{3}\pi + \Theta) J_{\frac{1}{3}}(\zeta) + \cos(\tfrac{1}{3}\pi - \Theta) J_{-\frac{1}{3}}(\zeta) \}, \quad (28)$$

with $\zeta = \int_0^z Q dz$ and Θ an arbitrary constant on the range $-\frac{1}{2}\pi < 0 \leq \frac{1}{2}\pi$ for $z \geq 0$.

When there is no other turning point for $z < 0$ (as happens in the present case) then $\Theta = 0$. The solution of equation (22) is then

$$\phi \sim A(\zeta/Q)^{\frac{1}{2}} \{ J_{\frac{1}{3}}(\zeta) + J_{-\frac{1}{3}}(\zeta) \}, \quad (29)$$

with

$$\zeta = \frac{1}{\alpha H_0} \left[(-aq^2 + bq - \tfrac{1}{4})^{\frac{1}{2}} - \frac{b}{2\sqrt{a}} \sin^{-1} \left(\frac{b - 2aq}{\sqrt{(b^2 - a)}} \right) - \tfrac{1}{2} \sin^{-1} \left(\frac{bq - \frac{1}{2}}{q\sqrt{(b^2 - a)}} \right) + \frac{\pi}{4} \left(1 - \frac{b}{2\sqrt{a}} \right) \right], \quad (30)$$

and

$$q \equiv \exp(\alpha H_0 z).$$

The explicit solution of equation (9), when wind shear is included, involves considerable mathematical complications. When the height variation of H is neglected, then a solution can be obtained which, when stable waves are possible, involves Bessel functions of imaginary order and argument. With some approximations the temperature gradient can be retained, to give a solution in Whittaker functions. Neither of these solutions will be further discussed, a knowledge of the conditions of bounding in the presence of wind shear being sufficient for present purposes.

5. APPLICATION TO THE MICROBAROMETRIC OSCILLATIONS

It appears to have been generally accepted for many years that the microbarometric oscillations were a manifestation accompanying Helmholtz (1889) waves occurring at a surface of discontinuity of wind and temperature in the atmosphere. The early history of the subject is summarized by Goldie (1925), who describes the contributions of Kelvin, Lamb and Schmidt. Goldie further developed the theory of such waves and applied it to a few examples of such pressure waves recorded crossing the British Isles. In the absence of fuller upper-air data it was not possible to test the theory adequately. A different light was thrown on the phenomena by the investiga-

tions of Johnson (1929), who found that the period of the observed oscillations agreed with that deduced by Brunt (1927) for the free period of vertical oscillation of a small element of atmosphere. Brunt's period corresponds in our terminology to

$$\sigma_B = \sqrt{\left\{ \frac{g}{H} \left(\frac{\gamma - 1}{\gamma} + \alpha H_0 \right) \right\}}$$

and is *the period corresponding to the horizontal part of curve II in figure 1*. In the discussion of Johnson's paper (1929) it was pointed out by Brunt and others that no sound theoretical reason had emerged why the observed period of the oscillations should agree with that of a *small element* of atmosphere. A further theoretical attack on the problem has been made recently by Pekeris (1948), who examines the propagation of a pressure pulse in a model atmosphere consisting of a troposphere of uniform lapse rate, capped by a uniform stratosphere. He finds that a pressure impulse of uniform spectrum from a point source at the ground tends mainly to excite the modes of atmospheric oscillation near to Brunt's period. It is difficult to see, however, what bearing this can have on the production of microbarometric oscillations, which are of common occurrence. A vital objection to Pekeris's theory seems to emerge from consideration of the velocities of horizontal travel of the disturbances. According to his theory they should have velocities only about 10% slower than sound waves at ground level; actually the available evidence shows that they travel at about 12 m./sec.

According to the view here presented the microbarometric oscillations are cellular waves travelling horizontally in a duct bounded below by the ground and on top by a discontinuity in temperature or temperature *gradient*. The period of the oscillations is confined to a narrow range by two boundary conditions, that oscillations must be possible in the lower stratum (1) right down to ground level ($Q_1^2 > 0$), while they must be impossible in the upper stratum (2) ($Q_2^2 < 0$). The waves are almost certainly produced by wind shear either in the duct itself, or, more probably, at the discontinuity at the upper boundary. An essential difference between the type of wave considered and Helmholtz waves lies in the fact that the latter are centred about the surface of discontinuity, and decay with distance from this surface, whereas the cellular waves occupy the volume between the two bounding surfaces and are rotational in type.

Simple types of temperature distribution capable of forming a bounded duct are shown in figure 2 (*A, B, C, D*). In *A* a ground inversion gives way at *O* to a normal lapse rate. In *B* the lapse rate *increases* at *O*. In *C* there is a discontinuity of temperature at *OO'*, the temperature at *O'* being higher than that at the ground, while *D* is a more physically likely version of *C*, with discontinuities of gradient at *O* and *O'*. In *A* and *B* the upper boundary of the duct is provided by the increased lapse rate above *O*. In *C* and *D* the boundary is occasioned by the temperature at *O'*, which is higher than that at the ground. From the discussion in § 3 above it will be realized that an adequate boundary is more readily formed by a small increase in lapse rate than by a small increase in temperature, so that *A* and *B* are more likely to be associated with the production of microbarometric oscillations than are *C* or *D*. (Johnson noted three examples of these oscillations accompanied by low-level inversions out of ten cases studied.)

For quantitative discussion, the type of temperature distribution shown in figure 2*B* is selected, the lapse rate being $\frac{2}{3}\Gamma$ below O and $\frac{4}{5}\Gamma$ above. Attention is confined to moderately short wave-lengths, of the order of a few km., such as are associated with the microbarometric oscillations. This means, in effect, that we are confined to the horizontal part of the bounding curve II in figure 1. It then follows from the equation for this curve that the possible values of σ are confined between 8×10^{-3} and 10.6×10^{-3} sec., i.e. roughly between periods of 10 and 13 min. Waves of period less than 10 min. would have their lower boundary above the ground, and so could not affect the barometer, while those of period longer than 13 min. would be unbounded on the upper surface, and their energy would be dissipated upwards.

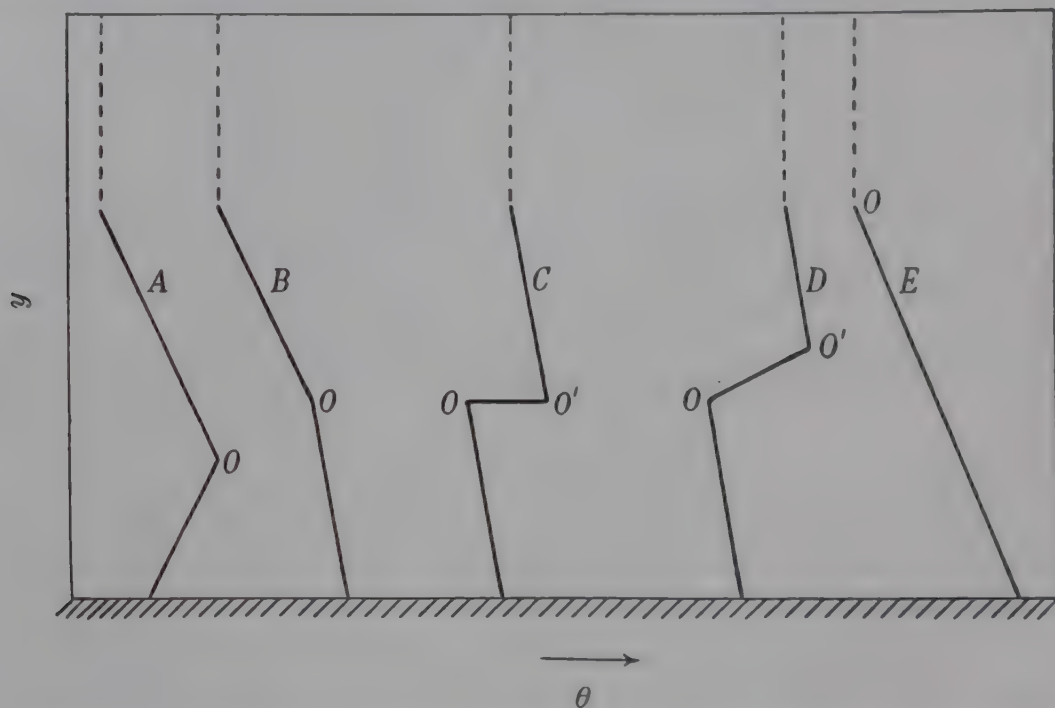


FIGURE 2. Discontinuities in lapse rate forming atmospheric wave ducts.

At O , ($z = h$) the condition for a free surface ($\phi = 0$) gives, by equation (25),

$$AJ_{\frac{1}{2}}\left(\frac{2}{3n}\eta_h^{\frac{1}{2}}\right) + BJ_{-\frac{1}{2}}\left(\frac{2}{3n}\eta_h^{\frac{1}{2}}\right) = 0. \quad (31)$$

At ground level ($z = 0$) the boundary condition $v = 0$ gives, by equation (27),

$$\begin{aligned} \eta_0^{\frac{1}{2}}(\sigma^2 - 2gk^2H_0) \left\{ AJ_{\frac{1}{2}}\left(\frac{2}{3n}\eta_0^{\frac{1}{2}}\right) + BJ_{-\frac{1}{2}}\left(\frac{2}{3n}\eta_0^{\frac{1}{2}}\right) \right\} \\ = \frac{2\sigma^2\eta_0}{n} \left\{ AJ_{-\frac{1}{2}}\left(\frac{2}{3n}\eta_0^{\frac{1}{2}}\right) - BJ_{\frac{1}{2}}\left(\frac{2}{3n}\eta_0^{\frac{1}{2}}\right) \right\}. \end{aligned} \quad (32)$$

Eliminating the arbitrary constants in the usual manner the relation between σ and k is obtained. Selected values of σ in the permitted range were then taken, and the equations solved for h for representative values of k . The results (for the fundamental modes) are shown in figure 3 in the form of curves for h (km.) plotted against wave-length (km.) for a series of values of σ . From these curves the σ, k relation is plotted for a series of values of h (figure 4). The curves in this latter figure give a complete

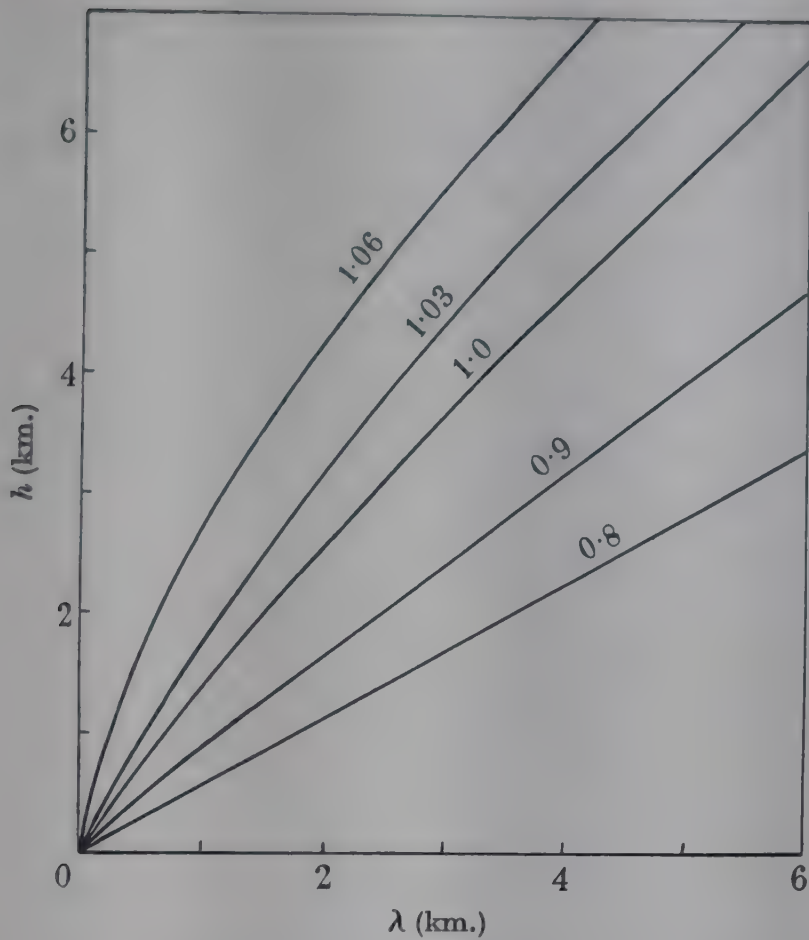


FIGURE 3. Relation between duct height and wave-length for various permissible frequencies, lapse rate $\frac{2}{3}$ adiabatic, fundamental mode. The numbers on the curves are the values of $10^2 \sigma$.

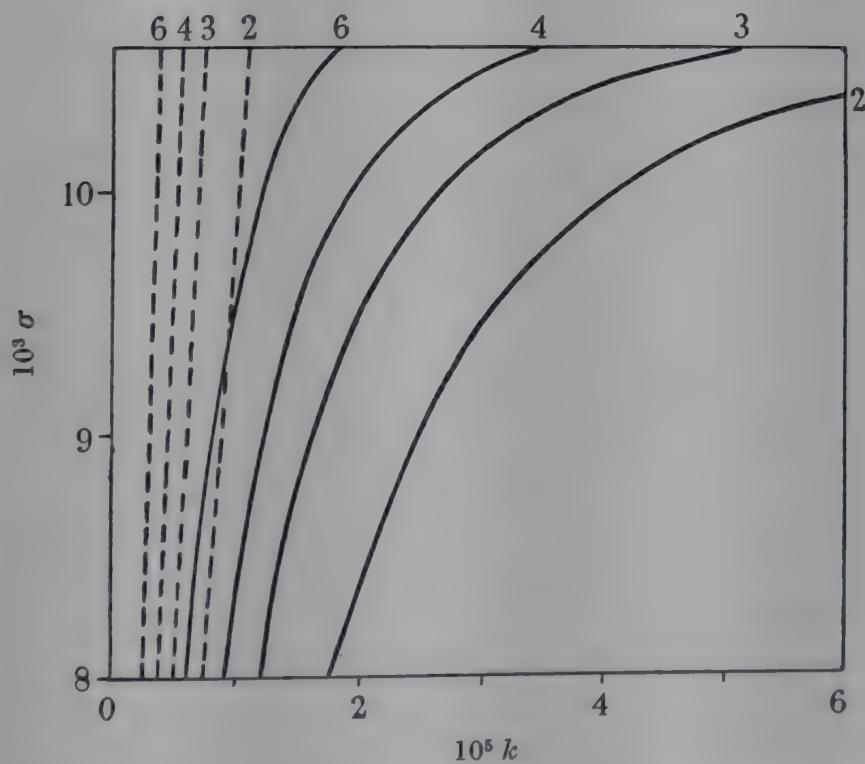


FIGURE 4. Dispersion relations for ducts of height 2, 3, 4, 6 km. respectively. Full curves are for lapse rate $\frac{2}{3}$ adiabatic, dashed curves for isothermal atmosphere.

description (for the fundamental mode) of the types of waves which can travel horizontally along ducts of various widths. Thus for a given value of duct width (h_1), and selected frequency (σ_1), the permissible value of k is read from the curve ($h = h_1$). The phase velocity of the wave is then σ_1/k_1 , and the group velocity ($d\sigma/dk$) is given by the slope of this curve at the point σ_1, k_1 . It is now possible further to narrow the range of physically permissible waves in a given duct. At the higher frequencies the dispersion is marked ($d\sigma/dk < \sigma/k$). At these frequencies the velocity of energy flow lags behind the wave velocity. Travelling waves of these frequencies are physically unlikely. At the lower frequencies permissible the phase and group velocities become substantially equal. It seems likely therefore, that the energy of excitation will go into the production of the lowest frequency which can be effectively trapped. This means that the frequency of observed microbarometric oscillations ought to be determined principally by the lapse rate in the upper medium, and should be only slightly higher than Brunt's frequency for this medium at the boundary level O . Johnson's data are not sufficiently accurate or numerous to test this refinement.

It is seen from figure 3 that the wave velocities at the lower permissible frequencies range from about 5 m./sec. in the shallower ducts (2 km.) to about 14 m./sec. in the deeper ducts (6 km.). These values seem to agree satisfactorily with the wave velocities and probable duct widths deducible from Goldie's (1925) data. In fact, the interesting possibility emerges of determining the lapse rate in the upper air and the height of the first discontinuity in this rate, from ground observations with three microbarographs spaced a few km. apart. With this disposition the wave-length and frequency of the oscillations would be determinable, these two quantities being sufficient to give the height of the duct and the lapse rate at its upper boundary, provided the lapse rate in the duct were known approximately.

It is now possible to see more clearly why the type of oscillation discussed by Pekeris (1948) seems inapplicable to the microbarometric oscillations. The temperature distribution discussed by him is shown in figure 2 *E*. It is necessary to find a type of oscillation which will be bounded by the stratosphere. Figure 1 shows that (with $\alpha = 0$) this can occur, for low frequencies, only on the sloping region of curve II (*a*). On this part of the curve the phase velocity tends to the value $2\{gH(\gamma - 1)/\gamma\}^{\frac{1}{2}}$ or $0.9c$, which is some 20 times the observed velocities.

6. TRAVELLING IONOSPHERIC DISTURBANCES

A typical disturbance in the ionosphere travels with a speed of about 10 km./min. Its wave-length is some hundreds of kilometres, and the apparent period of the order of half an hour. The disturbance consists of a rhythmic change of the electron densities in the parts of the F_2 region accessible to observation by radio-sounding techniques. The larger disturbances tend to have the longer wave-lengths and periods, and frequently consist of only one complete cycle of oscillation. One of the earliest studies of these disturbances appears to be that of Pierce & Mimno (1940) who built up by deduction, from comparatively slender evidence obtained at one site, a description which is remarkably close to that now obtained in Australia from simultaneous observations at several locations. Later, Wells, Watts & George (1946)

obtained evidence of what appeared to them to be vertical travelling 'ionization clouds' in the F_2 region. Much data of a similar nature has been compiled on cine-films by the Bureau of Standards, and circulated privately to interested laboratories. From the discussion below it will be seen that there is reason to believe that these apparently vertically travelling disturbances are identical with the horizontally travelling disturbances investigated by Pierce & Mimno, and by Munro in Australia. It will appear that the phase of the horizontally travelling disturbances usually advances with height in the F_2 region, thus giving the impression, when observed at a single site, of a downward progression of disturbance. On this view it would appear that all the phenomena reported above are of a single type, horizontally travelling pulses or waves, whose phase varies monotonically with height.

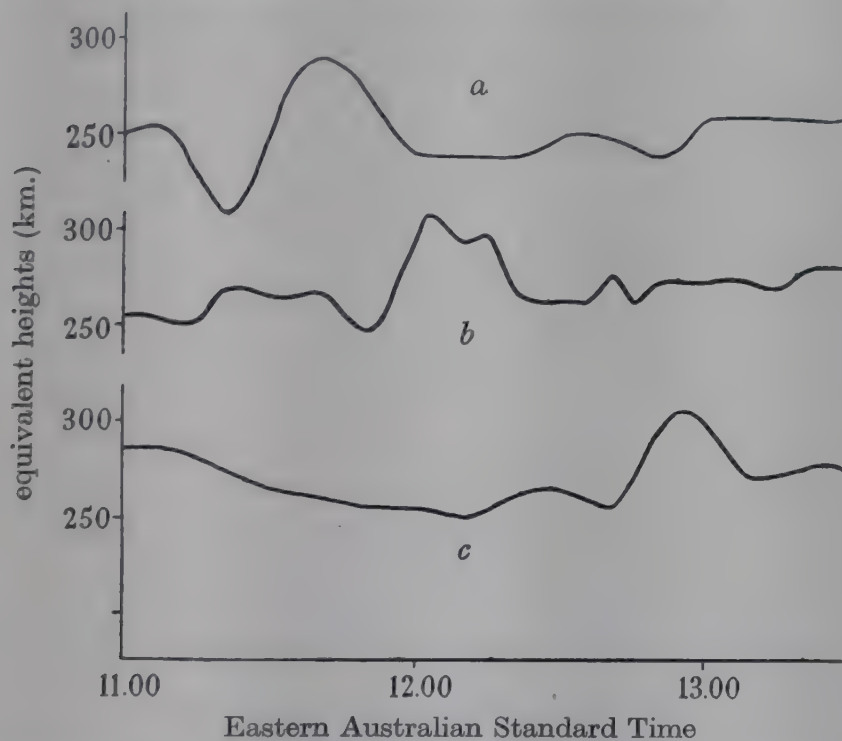


FIGURE 5. Ionospheric (F_2 region) disturbance passing over (a) Canberra, (b) Sydney and (c) Brisbane on 9 June 1948. Equivalent heights (km.) at a frequency of 6 Mc./sec. are plotted against Eastern Australian Standard Time.

In figure 5 is shown the passage of such a disturbance successively over Canberra, Sydney and Brisbane. Depicted is the equivalent height of the F_2 region, as given by the 'ordinary' reflected echo at a frequency of 6.0 Mc./sec. In this case the oscillation is maintained over one cycle, and it will be noticed that it is somewhat different in shape at each location. This is not surprising when one considers the dispersive properties of the medium, but it makes accurate timing difficult. In the example illustrated it appears that the disturbance arrived at Sydney 27 min. after reaching Canberra, and at Brisbane after another 49 min. This is consistent with a wavefront travelling E.N.E. with a speed of about 10 km./min. It is only the largest disturbances which can be tracked over such large distances (nearly 1000 km.), but smaller ones can be followed for several hundred kilometres.

A special study has been made of the disturbance of 22 June 1948 reported by Munro (1949) and analyzed by him in detail as it crossed Sydney. The records ob-

tained at Canberra at 10 min. intervals by the automatic P' - f echo-sounding equipment were analyzed in detail by the method due to Manning (1947). The analysis gives the true height and density gradient of ionization in the F_2 region at each ten minute interval between 11.50 and 12.50 hr. (Eastern Australian Standard Time). The results are plotted in figure 6, the six curves in which show the time variation of the heights at which six selected electron densities (1.1×10^6 to 1.6×10^6 electrons/cm.³) occurred. The oscillation in the heights at which these densities occurred can be clearly seen, as well as the systematic advance of phase of the disturbance as one passes upwards in the F_2 region. In these respects this disturbance is quite typical of many others observed.

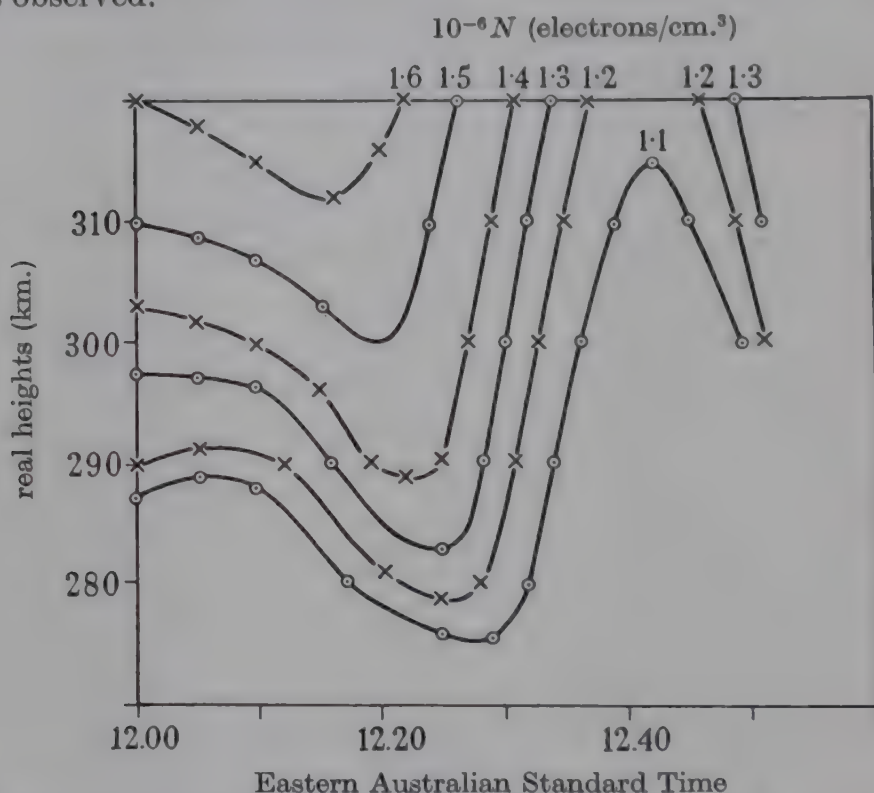


FIGURE 6. Height movements of six selected electron densities ($1.6, 1.5, \dots 1.1 \times 10^6$ electrons/cm.³) accompanying passage of F_2 region disturbance over Canberra on 22 June 1948.

7. THEORETICAL CONSIDERATIONS IN THE F_2 REGION OF THE IONOSPHERE

First it is desirable to discuss briefly certain salient ionospheric characteristics which must influence atmospheric oscillations at these levels. An outstanding fact is the high kinematic viscosity ($\nu = \mu/\rho$) in the ionosphere. The density (ρ) is approximately $1/(2 \times 10^8)$ of that at the ground, so that ν is approximately 10^7 c.g.s. units in the F_2 region. (This is comparable with the viscosity of treacle.) An important consequence must be a much-increased stability of the air for laminar flow. Prandtl (1932) has shown that stability in a non-viscous atmosphere depends on the dimensionless number $\sigma_B/\beta U_0$. When this number is less than about $\frac{1}{2}$ turbulence sets in. It is quite possible that the high gradients of wind velocity in the ionosphere would permit this number to be less than $\frac{1}{2}$, but the high viscosity should prevent the full development of turbulence. Instead, it seems likely that local perturbations in the stream line flow may develop. As will be seen later, it is likely that such perturbations are the cause of the observed disturbances.

Again it might be anticipated that high viscosity would damp out a travelling pressure wave. However, for the low frequencies under consideration this does not happen. Thus if it be crudely assumed for simplicity that the law of damping is similar to that for sound waves, then the waves would decay by $1/e$ in a distance l , where $l = \frac{3}{2}c^3/\nu\sigma^2$. This is of the order of 10^{12} cm. or more for the low frequencies considered, so that it can safely be assumed that viscous damping of these waves is negligible even in the F_2 region. There remains, however, inductive drag, the importance of which was first stressed by Cowling (1945). A mass of conducting air set in motion across the earth's magnetic field would lose its momentum in a time which is of the order of half an hour for the F_2 region. Since the observed disturbances can travel for some hours relatively undamped it would appear that forces must be continually at work maintaining them.

A factor of great importance in determining the electron motions consequent upon a given atmospheric motion is the earth's magnetic field. At the low pressures existing in the F_2 region electrons (and ions) are almost completely unable to move translationally in any direction other than along the lines of magnetic force (Martyn 1947). This means, for example, that while the atmospheric gases may move great distances as winds, the ionization produced in a given locality spends its lifetime near that locality, able to execute movements or oscillations only along the lines of force. It will be found below that this almost inescapable conclusion provides a natural explanation of the observed height gradient in the phase of the travelling disturbances.

Turning now to the bounding conditions of cellular atmospheric oscillations (equation (19)), it is apparent that oscillations of period comparable with 1 hr., in a region where H_0 is of order 40 km., can be bounded only if the lapse rate closely approaches the adiabatic. If γ retains its usual value of 1.4 in these regions this demands a lapse rate approaching 10° C/km. This is very unlikely, even for a limited range of heights at these levels. It is contrary to what would be expected from considerations of radiative equilibrium, and, if it extended over an appreciable range of heights, would reduce the temperature so much that there could scarcely be enough atmosphere left at 1000 km. to produce the sunlit aurorae.

A possible way out of this difficulty is to consider the conditions of bounding in the presence of horizontal winds, and when $\Omega \rightarrow 0$. This gives in effect a non-oscillatory convection (or Bénard) cell, which is carried along by the wind. In fact, all the available observational evidence, which will be discussed in a later communication, suggests that this type of cell, which occurs, for example, in billow clouds, is that responsible for the observed effects in the F_2 region. However, Sekera (1948) has shown, in his study of billow-cloud theory, that such cells too require a fully adiabatic or even superadiabatic lapse rate for bounding. Since high lapse rates in the F_2 region are unacceptable, there appears to be little chance of escaping the conclusion that γ in the F_2 region is considerably less than 1.4. There is reason to believe that this can occur at a level in the atmosphere where one of the principal constituents (viz. nitrogen) is beginning to dissociate into its atomic constituents. The principal condition necessary for reduced γ is that the total internal energy of the gas shall be considerably greater than the translational energy, but it is also necessary that the times

involved in the dissociation and recombinative processes shall be comparable with, or less than, the time to circulate the convection cell. The first condition can readily be satisfied, but careful investigation, which cannot be attempted here, will be necessary before it is possible to say that the second condition is possible in the F_2 region. However, whatever the theoretical implications, the observational fact that F_2 disturbances can travel 1000 km. relatively undamped makes it very difficult to avoid the conclusion that they are adequately bounded, and such bounding seems to necessitate a value for γ nearer to 1.0 and 1.4.*

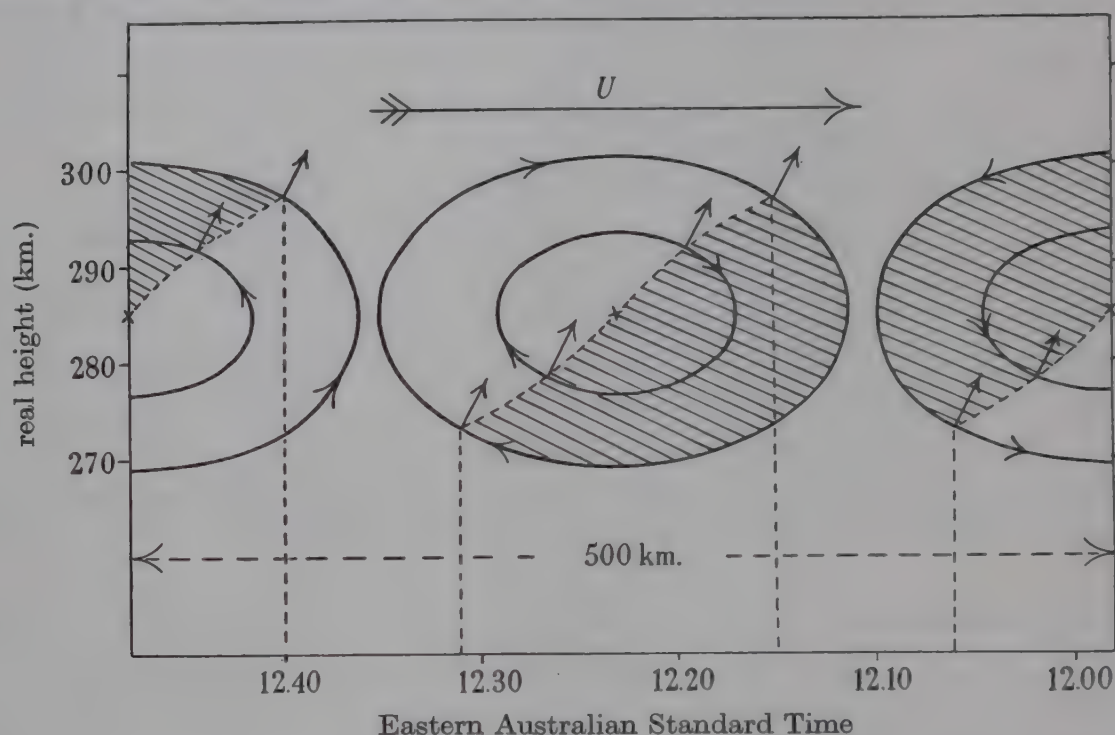


FIGURE 7. Travelling atmospheric cells which semi-quantitatively account for the electron movements of figure 6. Cells travel with the wind (U). Air motion in cells follows arrow-heads. Short arrows show inclination of magnetic field (60°).

It is now possible to illustrate in some detail how a typical travelling atmospheric cell can produce the observed F_2 disturbance of figure 6 in its passage. For illustrative purposes convection cells ($\Omega = 0$) are shown (in figure 7) travelling from left to right with the wind. The time scale (at Canberra) therefore runs from right to left. The cells are drawn as ellipses, which is a sufficient approximation for present purposes. The scale is expanded some four times in the vertical direction, for clarity of presentation. The direction of the (perturbing) wind velocities in the cells is shown by arrows on the ellipses. The inclination of the magnetic field (60°) is shown by short arrows at those points on the ellipses where this field is perpendicular to the perturbed motion. At these points there is no tendency for the perturbing wind to drive electrons either up or down. In the shaded regions of the cells the perturbing wind drives electrons (and ions) downwards; in the unshaded portions, upwards. Comparison of this model with

* It is perhaps significant that another region of the ionosphere which can exhibit an erratic patchy structure (the sporadic E ionization) occurs at the height where molecular oxygen becomes markedly dissociated. At these heights (c. 100 km.) the kinematic viscosity of the air, and also the effects of inductive drag, are very much less than in the F_2 region, so that marked turbulence may develop if γ approaches 1.

figure 6 shows that a natural interpretation of the observed movements of ionization is obtained. Thus, at the 275 km. level, ionization ceases to move downwards at about 12.30 hr. Soon after this (at about 12.33) the lines of flow at all heights run approximately up along the field; there is therefore rapid upward flow at all heights, as figure 6 shows. Again, at 12.17 hr. there is rapid downward flow at all heights below 300 km., but at greater heights the ionization has begun to move up. Soon after 12.40 the ionization at heights of about 310 km. and upwards has begun to move down again. It is to be emphasized that figure 7 is illustrative and not quantitative. Many complications are ignored, such as the precise shape of the convection cells, and the effective inclination of the magnetic field when the wind is not due N.-S. (magnetic), while in interpreting the diagram the complications introduced by the convergence of ionization (due to height gradient of the perturbing wind) are similarly neglected. Nevertheless, it seems clear that the detailed quantitative interpretation of these ionospheric disturbances can, and possibly must, be fitted into the pattern of cellular movement broadly outlined.

The particular disturbance here examined is typical in its morphology, but is unusually large in amplitude. It was for this reason that it has been selected for special study here. Its unusual amplitude is accompanied by a wave-length and period which are both about twice the average: in all other respects it is typical. It will be seen that the centre of the cells are at a height of about 290 km. It would not be safe, however, to draw the conclusion that this height is preferred by travelling cells. If these cells travelled above the level of maximum ionization in the F_2 region they would not be observed by radio methods, while if they travelled much lower they would enter a region where the time of recombination of electrons is comparable with, or less than, the time to complete the circuit of the cell, and they would therefore produce only small *ionospheric* disturbances.

This work has been carried out for the Radio Research Board of the Commonwealth Scientific and Industrial Research Organization, Australia, and is published by permission of these bodies. The author is indebted to G. H. Munro for much discussion of his observations on travelling ionospheric disturbances.

REFERENCES

- Beynon, W. J. G. 1948 *Nature*, **162**, 887.
 Brunt, D. 1927 *Quart. J. R. Met. Soc.* **53**, 30.
 Cowling, T. G. 1945 *Proc. Roy. Soc. A*, **183**, 453.
 Dinnik, A. 1911 *Arch. Math. Phys., Lpz.*, **18**, 337.
 Goldie, A. H. R. 1925 *Quart. J. R. Met. Soc.* **51**, 239.
 Helmholtz, H. v. 1889 *S.B. preuss. Akad. Wiss.* p. 761.
 Johnson, N. K. 1929 *Quart. J. R. Met. Soc.* **55**, 19.
 Lamb, H. 1908 *Proc. Lond. Math. Soc.* **7**, 122.
 Langer, R. E. 1937 *Phys. Rev.* **51**, 669.
 Manning, L. A. 1947 *Proc. Inst. Radio Engrs, N.Y.*, **35**, 1203.
 Martyn, D. F. 1947 *Proc. Roy. Soc. A*, **189**, 241.
 Martyn, D. F. 1948 *Proc. Roy. Soc. A*, **194**, 445.
 Martyn, D. F. & Pulley, O. O. 1936 *Proc. Roy. Soc. A*, **154**, 455.
 Munro, G. H. 1948 *Nature*, **162**, 886.

- Munro, G. H. 1949 *Nature*, **163**, 813.
 Olivier, C. P. 1947 *Proc. Amer. Phil. Soc.* **91**, 315.
 Pekeris, C. L. 1948 *Phys. Rev.* **13**, 145.
 Pierce, J. A. & Mimno, H. R. 1940 *Phys. Rev.* **57**, 95.
 Prandtl, L. 1932 *Beitr. Phys. frei Atmos.* **19**, 188.
 Scorer, R. S. 1949 *Quart. J. R. Met. Soc.* **75**, 41.
 Sekera, Z. 1948 *J. Met.* **5**, 93.
 Watson, G. N. 1944 *Theory of Bessel functions*. Cambridge University Press.
 Wells, H. W., Watts, J. M. & George, D. E. 1946 *Phys. Rev.* **69**, 540.

Unified field theory in six dimensions

BY J. PODOLANSKI, *University of Manchester*

(Communicated by R. E. Peierls, F.R.S.—Received 12 October 1949)

The geometry of the Dirac equation is actually six-dimensional. Taking this hint it is shown that a six-dimensional—classical—field theory avoids the difficulties with which the Kaluza-Klein theory has to contend. Moreover, the possibility is gained of making the field energy of a point source finite.

A geometric requirement (structure axiom) decomposes the six-dimensional continuum. The space-time world appears as a subspace, immersed in the six-dimensional space. Each world point corresponds to a sheet of physically indistinguishable points. Potentials and gauge transformation can be interpreted geometrically.

As a consequence of the embedding, *two* fields of inertial forces (fields of constraint) occur. The one behaves just as the Maxwell field; the other has negative energy and can cancel the singularities of the electromagnetic one.

The theory can be made conformally invariant and then yields a relation between the mass and the charges (the sources of the constraining fields). The presentation given in this paper is provisional, as it gives infinite range also to the field of negative energy.

1. INTRODUCTION

Why six dimensions?

This attempt to combine gravitation and electromagnetism in a geometrical frame is an extension of the Kaluza-Klein theory (Kaluza 1921; Klein 1926, 1927). The latter shows some fascinating aspects: for example, the convincing unification of the conservation laws, the interpretation of the gauge transformation. Nevertheless, this theory has turned out to be sterile; its formalism is too vague to give proper guidance. The more precise formulation, which it received in the projective version (Schouten 1935), showed up its shortcomings even more clearly. Besides, this theory is ambiguous (with respect to the choice of the connexion), it cannot define a line element in the five-dimensional space and therefore not, without some arbitrariness, a geodesic. Finally, the formulation of the five-dimensional laws of nature becomes a toilsome and unpleasant task. The heuristic value of such a theory is nil.

Still there are reasons to believe that a hyperdimensional description of nature is useful. Although the theory to be considered is classical, the strongest argument is derived from the structure of the Dirac matrices. Some of these are associated with the space-time world, representing its $\binom{4}{2}$ possible infinitesimal rotations. However,

not only these six, but all the fifteen Dirac matrices are rotation operators.* The complete set represents the $\binom{6}{2}$ rotations in a six-dimensional space,† which includes the space-time world as a subspace.

This very picture—the world as a subspace of a six-dimensional manifold—is the one on which this suggestion is based. I shall try to show that this conception leads to a description which geometrizes both gravitation and electromagnetism, making them inertial forces.

But even more: the electromagnetic field appears to be accompanied by yet another field, the energy of which is *negative*. So it seems that this theory may contribute to the problem of self-energy. Compared with other attempts in that direction (Bopp, Stückelberg, Pais, Feynman) this theory has the advantages that the equations of the second field need not be guessed, that the relativistic invariance looks after itself, and that the negative sign of the energy appears naturally. Finally, we have reason to expect that the difficulties connected with the concept of fields of negative energy will not arise in this theory.

How to get rid of two dimensions

The first task of a theory of this kind is, of course, to ensure that the ‘events’ in the six-dimensional geometry take place in four dimensions, i.e. that the six-dimensional geometry gives a four-dimensional physics. In the Kaluza-Klein theory the corresponding step was taken by a projective interpretation of the co-ordinates and by conditions of homogeneity. We do not believe that the projective interpretation‡ is useful, and we replace the homogeneity conditions by a geometrical axiom that imposes certain restrictions on the connexion of the six-dimensional space. As a consequence of this ‘structure axiom’ the space becomes laminated, folds up into a four-parametric family of two-dimensional sheets. The points on a sheet turn out to be physically indistinguishable; each sheet corresponds to *one* point in the space-time world. The physical events take place in the four-dimensional spaces normal to the sheets. Therefore this attempt may be classified as an embedding theory, the electromagnetic forces having the character of forces of constraint.

The structure axiom only tells us that the space is laminated. How it is built up from the layers is determined by the field equations, which give the connexion between the structure and the material sources (charges). This process is similar to that of general relativity. The latter postulates that the space is Riemannian, the metric tensor itself being given by the masses.

As our structure axiom we suppose the space to admit of a two-dimensional translation. For this assumption no reason can be stated. Indeed, it will afterwards be shown that this condition is sufficient to give to the resulting physics features with which we are familiar. Thus the conservation of charge and mass is a consequence of

* They satisfy the structure relations of the rotation group.

† Its metric is not arbitrary, if the Dirac matrices have a hermiteizing matrix. The space has two time-like dimensions (Schouten & Haantjes 1935); thus the two-dimensional hyper-space has an indefinite metric.

‡ In this direction other attempts have been made (Hoffmann 1947, 1948) which seem to have little connexion with ours.

this axiom. We cannot exclude the possibility that the structure axiom may only express a first approximation, equivalent to familiar physics. But, as the axiom surely is connected with the fixing of a scale of length, we hope that a conformally invariant version of this theory will throw some light on its significance.

Geometrically the axiom means a restriction imposed on the connexion of the six-dimensional space. While other unitary theories seek the solution in a generalization of the four-dimensional connexion, we find it in a specialization of a Riemannian connexion—but of a six-dimensional space. The effect is similar: the connexion, induced in the four-dimensional subspace, shows deviations from a Riemannian. The ‘fields of constraint’ are an expression for these deviations.

How to do physics in six dimensions

We get the equations of motion by assigning to a mass point for its path a six-dimensional *geodesic*. The four-dimensional observer perceives a pair of inertial forces (constraining forces, embedding forces) which have the form of the Lorentz force.

The field equations are derived from a variation principle the Lagrangian of which is the six-dimensional *curvature*. Then both constraining fields satisfy the Maxwell equations. This result is not satisfactory: to have a physical meaning, the second of these fields—of negative energy—ought to have a finite range. Indeed, in a more rigorous treatment, we obtain for this field Proca’s equations.

This paper thus does not give ‘the whole truth’. Here is offered a simplified theory which allows us to explain the essential points—without leaving the framework of classical physics. The introduction of fields with finite range is essentially a quantum-mechanical step—at least in this attempt—and will be carried out in a subsequent paper.

STRUCTURE AXIOM

2. MOTION

(a) *Infinitesimal transformation*. We start from a Riemannian space, described by the co-ordinates ξ^λ , the metric tensor $g_{\lambda\mu}$ and the linear connexion, based upon the latter.* For the covariant derivative we write $\hat{\nabla}_\lambda$:

$$\hat{\nabla}_\lambda v_\mu = \partial_\lambda v_\mu - \hat{\Gamma}_{\lambda\mu}^\nu v_\nu,$$

$$\partial_\lambda = \frac{\partial}{\partial \xi^\lambda}; \quad \hat{\Gamma}_{\lambda\mu}^\nu = \left\{ \begin{matrix} \nu \\ \lambda\mu \end{matrix} \right\}.$$

This space may undergo a transformation, given by the vector field p^λ :

$$\delta \xi^\lambda = \delta t \cdot p^\lambda,$$

where δt is a constant infinitesimal scalar. Then $d\xi^\lambda/dt = p^\lambda$ is the differential equation of the paths of the transformation; their tangential vector p^λ will be called the *path vector*. The transformation, which a geometric entity X experiences,

$$\delta X = \delta t \cdot D_{(1)} X,$$

* $\lambda, \mu = 1, \dots, n$, later on $1, \dots, 6$.

is determined by the operator $D_{(1)}$ which depends upon the transformation properties of X . For a scalar u of course

$$D_{(1)}u = p^\rho \partial_\rho u = \partial_{(1)}u,$$

where $\partial_{(1)} = p^\rho \partial_\rho$ is the derivative in the path direction. Moreover, for the components of a vector, the deformation of the co-ordinate system has to be taken into account,

$$D_{(1)}v_\lambda = \partial_{(1)}v_\lambda + v_\rho \partial_\lambda p^\rho.$$

On the right-hand side we may replace the derivatives ∂_λ by the covariant derivatives $\hat{\nabla}_\lambda$,

$$D_{(1)}v_\lambda = \hat{\nabla}_{(1)}v_\lambda + p_\lambda{}^\rho v_\rho, \quad (1)$$

with

$$p_{\lambda\mu} = \hat{\nabla}_\lambda p_\mu.$$

This time, $\hat{\nabla}_{(1)} = p^\rho \hat{\nabla}_\rho$ gives the covariant derivative in the direction of the path. Thus, compared with the vector displaced *parallel* by $\delta\xi^\lambda = \delta t \cdot p^\lambda$, the vector v_λ undergoes a transformation described by the affiner $\delta t \cdot p_{\lambda\mu}$. The symmetric part $p_{(\lambda\mu)} = \frac{1}{2}(p_{\lambda\mu} + p_{\mu\lambda})$ gives the deformation of the vector field, the skew-symmetric part $p_{[\lambda\mu]} = \frac{1}{2}(p_{\lambda\mu} - p_{\mu\lambda})$ the rotation, which v_λ undergoes in the course of the displacement.

(b) *Motion*. A transformation which does not affect lengths and angles is called a motion. For this it is necessary that the metric tensor remain unchanged,

$$D_{(1)}g_{\lambda\mu} = 0.$$

From (1) this is equivalent to

$$p_{(\lambda\mu)} = \hat{\nabla}_{(\lambda} p_{\mu)} = 0.$$

This condition, called Killing's equation, limits the transformation to rotations, the affiner

$$p_{\lambda\mu} = \hat{\nabla}_\lambda p_\mu = \partial_{[\lambda} p_{\mu]}$$

is skew-symmetric.

The existence of a vector field p^λ , for which $\hat{\nabla}_{(\lambda} p_{\mu)} = 0$, is not compatible with an arbitrary Riemannian connexion. The requirement that the space admit of a motion imposes restrictions upon the connexion. Such spaces have the important property that the equation of their geodesics has an integral of the *first* order. Given u^λ as the tangential vector of the geodesic, then its differential equation is (since the geodesic arises by a parallel displacement of the tangential vector in its own direction)

$$u^\rho \hat{\nabla}_\rho u^\lambda = 0.$$

This equation always has an integral of the second order

$$u^\lambda u_\lambda = \text{const.}$$

In our case, however, we have, in addition, an integral of the first order

$$p^\lambda u_\lambda = \text{const.}$$

Indeed, the left-hand side is constant along the geodesic because of the skew-symmetry of $\hat{\nabla}_\mu p_\lambda$.

(c) *Translation*. Such is called a motion for which the path vector has a constant length

$$p^\rho p_\rho = \text{const.}$$

By differentiation one finds ($p_{\rho\lambda} = \hat{\nabla}_\rho p_\lambda$)

$$p^\rho p_{\rho\lambda} = 0.$$

The skew-symmetric affnor $p_{\lambda\mu}$ has a null direction, the affnor covers at the most an $(n-1)$ -dimensional domain.

A translation p^λ is thus defined by

$$\left. \begin{aligned} p^\rho p_\rho &= a^2 = \text{const.}, \\ D_{(1)} g_{\lambda\mu} &= 0. \end{aligned} \right\} \quad (2a)$$

For the field $\hat{\nabla}_\lambda p_\mu$ it follows

$$\left. \begin{aligned} p_{\lambda\mu} &= \hat{\nabla}_\lambda p_\mu = \partial_{[\lambda} p_{\mu]}, \\ p^\rho p_{\rho\lambda} &= 0, \\ D_{(1)} p_{\lambda\mu} &= 0. \end{aligned} \right\} \quad (2b)$$

The property that $p_{\lambda\mu}$ is a rotation (first equation (2b)) can also be expressed by

$$\partial_{[\lambda} p_{\mu\nu]} = 0.$$

The square brackets indicate again the skew-symmetric part, i.e. here

$$\partial_\lambda p_{\mu\nu} + \partial_\mu p_{\nu\lambda} + \partial_\nu p_{\lambda\mu} = 0.$$

It can easily be proved that $p_{\lambda\mu}$ is invariant with respect to translation (last equation (2b)). From the second equation one gathers that the paths of a translation are geodesic.

3. STRUCTURE AXIOM

We now formulate our structure axiom: *Space admits of a two-dimensional translation.*

(a) The translation has to be two-dimensional, to give account of the four-dimensionality of the space-time world. But really it is sufficient to postulate only a translation, in order to cause the jump from six dimensions to four. In fact, we saw a translation to be equivalent to the existence of a skew-symmetric affnor $p_{\lambda\mu}$ which has a null direction p^λ ,

$$p^\rho p_{\rho\lambda} = 0.$$

Since a skew-symmetric matrix has even rank, $p_{\lambda\mu}$ is four-dimensional and there is a second vector q^λ , so that also

$$q^\rho p_{\rho\lambda} = 0.$$

This vector can be assumed as normalized and perpendicular to p^λ . If we want to make the description in both the vector fields symmetrical, then q^λ has also to be a field of motion. This has as consequences:

(i) the affnors $p_{\lambda\mu}$ and $q_{\lambda\mu}$ cover the *same* four-dimensional domain, their common null space is generated by the path vectors p^λ, q^λ ;

(ii) the null domain of the two affnors is *holonomic* (V_2 -generating). In other words, the duple of the path vectors envelops a four-parametric family of two-dimensional *sheets*;

(iii) the sheets are flat and geodesic subspaces with indefinite metric.

As a consequence of the structure axiom the six-dimensional space is seen to be built up from a family of two-dimensional planes. These planes (sheets) are generated

by the paths of the two translations. They are the 'invariant varieties' of the motion. By a translation each point of a sheet is transformed into another point on the same sheet; a translation produces a motion of the sheets within themselves.

(b) *Proof.* The system (2a), expressing p^λ to be a translation, is accompanied by a similar system for q^λ :

$$\begin{aligned} p^\rho p_\rho &= a^2, & q^\rho q_\rho &= b^2, \\ D_{(1)} g_{\lambda\mu} &= 0, & D_{(2)} g_{\lambda\mu} &= 0. \end{aligned}$$

Here q^λ is defined as the second null direction of $p_{\lambda\mu} = \hat{\nabla}_\lambda p_\mu$ (the first being p^λ),

$$q^\rho p_{\rho\lambda} = 0.$$

From this and from $p^\rho q_\rho = 0$ follows

$$p^\rho q_{\rho\lambda} = 0,$$

i.e. the identity of the null spaces, which belong to $p_{\lambda\mu}$ and $q_{\lambda\mu}$. To prove the holonomy of the null space, we show that the commutator of the two derivatives $\partial_{(1)} = p^\rho \partial_\rho$ and $\partial_{(2)} = q^\rho \partial_\rho$ vanishes:

$$\begin{aligned} \partial_{[(1)} \partial_{(2)]} &= \frac{1}{2} (p^\rho \partial_\rho q^\lambda - q^\rho \partial_\rho p^\lambda) \partial_\lambda \\ &= \frac{1}{2} (p^\rho q_{\rho}{}^\lambda - q^\rho p_{\rho}{}^\lambda) \partial_\lambda = 0. \end{aligned}$$

The vectors p^λ , q^λ thus constitute rectangles of constant side length, and these form the sheets. The latter are geodesic as we gather from the group of equations $r^\rho \hat{\nabla}_\rho s^\lambda = 0$, where r^λ and s^λ stand for either vectors p^λ , q^λ .

(c) One of the path vectors p^λ , q^λ has to be space-like, the other time-like. We put therefore

$$p^\lambda p_\lambda = -q^\lambda q_\lambda = a^2 > 0.$$

We can write our formulae a little more economically, if we combine p^λ , q^λ in $r_{(s)}^\lambda$ with $s = 1, 2$. The square of their lengths and their product we can, on account of

$$p^\lambda p_\lambda = r_{(1)}^\lambda r_{(1)\lambda}^\mu g_{\lambda\mu} \quad \text{etc.},$$

write as $g_{(1)(1)}$, $g_{(2)(2)}$, $g_{(1)(2)}$. Then the conditions for a two-dimensional translation will be

$$\left. \begin{aligned} g_{(1)(1)} &= -g_{(2)(2)} = a^2; & g_{(1)(2)} &= 0, \\ D_{(s)} g_{\lambda\mu} &= 0 & (s = 1, 2). \end{aligned} \right\} \quad (3a)$$

Besides we have the condition of *holonomy* (the group condition for the translations), which we can put in the form

$$\partial_{[(s)} r_{(t)]}^\lambda = 0. \quad (3b)$$

The equations (3a, b) express the structure axiom. It gives the affinors* $\hat{\nabla}_\lambda r_{\mu}^{(t)}$ the following properties:

$$\left. \begin{aligned} r_{\lambda\mu}^{(t)} &= \hat{\nabla}_\lambda r_{\mu}^{(t)} = \partial_{[\lambda} r_{\mu]}^{(t)} \\ r_{(s)}^\rho r_{\rho\lambda}^{(t)} &= 0 \\ D_{(s)} r_{\lambda\mu}^{(t)} &= 0 \end{aligned} \right\} \quad (s \text{ and } t = 1, 2), \quad (3c)$$

* The quantity $r_{\lambda\mu}^{(t)}$ is a generalization of the 'second fundamental tensor' of differential geometry and is called (Schouten & Struik 1938) 'curvature affinor of the third order'. We shall refer to $r_{\lambda\mu}^{(t)}$ as the *embedding fields*.

i.e. the embedding fields are skew-symmetric, four-dimensional and invariant with respect to the translations $r_{(s)}^\lambda$.

The structure axiom therefore decomposes the space into a four-dimensional family of sheets which are flat and geodesic. The metric tensor $g_{\lambda\mu}$ and the embedding fields $r_{\lambda\mu}^{(n)}$ are invariant with respect to the translations along a sheet. The components of $g_{\lambda\mu}$ along it are constant, those of $r_{\lambda\mu}^{(n)}$ disappear. The embedding fields are rotations, therefore satisfy

$$\partial_{[\lambda} r_{\mu\nu]}^{(n)} = 0$$

or

$$\partial_\lambda r_{\mu\nu}^{(n)} + \partial_\mu r_{\nu\lambda}^{(n)} + \partial_\nu r_{\lambda\mu}^{(n)} = 0.$$

The equation of the geodesic (tangential vector u^λ) now has *two* integrals of the first order

$$r_{(s)}^\lambda u_\lambda = u_{(s)} = \text{const.},$$

where thus $u_{(s)}$ is the projection of the tangential vector on to the sheet.

(d) *The normal spaces.* By introducing at each point in space four ($i = 1, \dots, 4$) vectors n_i^λ normal to the sheet, we make up the duple $r_{(s)}^\lambda$ of the path vectors to a sexuple. The quadruple n_i^λ of normal vectors can be chosen in such a way (§ 16 (a)) that for each geometric entity

$$D_{(s)} X = r_{(s)}^\lambda \partial_\lambda X = \partial_{(s)} X,$$

if X refers to the sexuple. Then the chief consequences of the structure axiom can be formulated as follows: ($i, k, l = 1, \dots, 4$; $s, t = 1, 2$):

- (i) $r_{ik}^{(n)}$ are the only non-vanishing components of the embedding fields;
- (ii) metric tensor and embedding fields are constant along a sheet

$$\partial_{(s)} g_{ik} = \partial_{(s)} r_{ik}^{(n)} = 0; \quad (4a)$$

- (iii) the embedding fields satisfy the first set of Maxwell equations

$$\partial_i r_{kl}^{(n)} + \partial_k r_{li}^{(n)} + \partial_l r_{ik}^{(n)} = 0, \quad (4b)$$

where $\partial_i = n_i^\lambda \partial_\lambda$ are the derivatives along the normals.

In contrast to the path vectors $r_{(s)}^\lambda$, the normal set n_i^λ has not the property of holonomy. The quadruple does not envelop a four-dimensional co-ordinate space. The sheets have no orthogonal trajectories. We shall discuss the anholonomy of the normal set (i.e. of the space-time world) in the last chapter. It is this property which we can make responsible for the appearance of fictitious forces.

(e) The definition of motion, upon which we based the structure axiom, is not invariant with respect to a rotation of the vectors p^λ, q^λ in the sheet. The axiom defines two vector fields, not a two-dimensional direction field. But we need only the latter to realize the decomposition of the space. This generalization can, however, only be understood in terms of quantum theory, and its consideration is therefore postponed to a later publication.

GEOMETRICAL AND PHYSICAL CONSEQUENCES

Now we are going to describe the consequences of the structure axiom, anticipating some results of the next chapters. The following notation is used:

- $\lambda, \mu, \nu, \dots$ run from 1 to 6 (co-ordinate system ξ^λ),
- $i, k, l, \dots$ run from 1 to 4 (quadruple n_i^λ ; world components),
- $(s), (t), (u), \dots$ have the values (1), (2) (duple $r_{(s)}^\lambda$; sheet components),
- $a, b, c, \dots$ have the six values 1, ..., 4, (1), (2) (sexuple $n_i^\lambda, r_{(s)}^\lambda$).

4. DECOMPOSITION OF SPACE

The condition 'space admits of a two-dimensional translation' imposes a structure on the six-dimensional space. It splits up into

- (i) the two-dimensional sheets, generated by the path vectors $r_{(s)}^\lambda$,
- (ii) the four-dimensional vector spaces, covered by the embedding affinors $r_{\lambda\mu}^{(t)} = \hat{V}_\lambda r_\mu^{(t)}$.

The two spaces are normal to each other.

(a) If we complete the system of reference by the quadruple n_i^λ ($i = 1, \dots, 4$) normal to the sheets, the components $r_{ik}^{(t)}$ are the only non-vanishing ones of the embedding fields. Similarly only the projection g_{ik} of the metric tensor on to the four-dimensional subspace is a geometrical entity: The other components vanish or are (disposable) constants. As metric and embedding together make physics, all physical events take place in the four-dimensional normal space, which we have to identify therefore with the space-time world.

(b) Along a sheet the affinors g_{ik} and $r_{ik}^{(t)}$ are constant:

$$\partial_{(s)} g_{ik} = \partial_{(s)} r_{ik}^{(t)} = 0.$$

Thus all ∞^2 local normal spaces of any sheet have the same geometric and physical properties. Points on the same sheet, i.e. those which can be transformed into one another by a translation, are indistinguishable. Each sheet corresponds to *one* point in the world.

A translation, i.e. a motion of the sheets within themselves, is not a physical event. It is, in fact, equivalent to a gauge transformation. The *gauge invariance* of a quantity X is therefore the same as the invariance with respect to a translation, and it is expressed by its property of being constant along a sheet,

$$\partial_{(s)} X = 0.$$

The structure axiom is consequently identical with the condition that metric tensor and embedding affinors are gauge invariant. Gauge invariance is attributed—in a classical theory—to all physical entities (like velocity, energy-momentum tensor). Only quantum theory gives non-gauge-invariant entities a meaning.

5. EQUATIONS OF MOTION

The sheets are in general not parallel. The embedding fields $r_{\lambda\mu}^{(s)}$ just tell us how two neighbouring sheets are twisted relative to each other. If we go in the direction $\delta\xi^\mu$ from one sheet to the other,

$$\delta r_\lambda^{(s)} = \delta\xi^\mu r_{\mu\lambda}^{(s)}$$

gives the rotation which the subtending path vectors undergo. Of course, the normal space, the world of the four-dimensional observer, is turning with it. Two vectors in different space-time points (i.e. on different sheets), which are 'in reality' (i.e. six-dimensionally) parallel, no longer appear parallel to the observer; a geodesic path is not geodesic to him. The path of a material point (we assume it a geodesic) appears curved to him the same amount, as 'in reality' his system of reference (the normal quadruple) is turned. This rotation, described by the fields $r_{ik}^{(s)}$, is thus seen by him as a curvature of the path, and he will interpret it as an effect due to forces. These fictitious forces have the form of the Lorentz force. In our picture they are pure inertial forces.

6. FIELD EQUATIONS

The *field* is a purely geometric entity, the six-dimensional metric tensor $g_{\lambda\mu}$. Corresponding with the decomposition of the vector space we could choose the determining quantities of the metric field in such a way that one set— g_{ik} —describes the metric of the space-time world, the other— $r_{ik}^{(s)}$ —its embedding.*

By choosing as *Lagrangian* the six-dimensional curvature scalar, the field equations for $g_{\lambda\mu}$ therefore split up into two systems:

For g_{ik} we get the gravitational equations of relativity.

Each of the two fields $r_{ik}^{(s)}$ satisfies the Maxwell equations. Yet they differ in one essential point: The sources of $r_{ik}^{(1)}$ repel each other, those of $r_{ik}^{(2)}$ attract each other, if they have the same sign. The first field has positive, the other negative energy. (This difference arises because we made $r_{(1)}^\lambda$ space-like, $r_{(2)}^\lambda$ time-like.)

The first field $r_{ik}^{(1)}$ we may identify with the electromagnetic field. To the other field $r_{ik}^{(2)}$ we can give a physical interpretation only if it has a finite range. Indeed, the above-mentioned weakening of the structure axiom gives for the second embedding field the Proca equations.

If there are no sources, $r_{\lambda\mu}^{(s)}$ disappears. With this the decomposition of the space—given by the null directions of these affinors—becomes indefinite. The empty space has no structure.

7. POTENTIALS

There exists another possibility of describing the distribution of the metric field between gravitation and electricity. This is obtained by going over from the decomposition of the vector space (the sexuple) to one of the co-ordinate space. As every sheet corresponds to *one* point of the world, the world co-ordinates have to be constant along a sheet: The sheets are the co-ordinate hypersurfaces $x^i = \text{const.}$ ($i = 1, \dots, 4$). On each sheet we have to draw a parameter net $w^{(1)}, w^{(2)}$, to complete the co-ordinate system.

Two tangential vectors $\partial\xi^\lambda/\partial w^{(s)}$ of this co-ordinate system lie in a sheet. But the other four $\partial\xi^\lambda/\partial x^i$ are in general not normal to the sheet and thus do not carry the space-time world. The resolution of a vector with respect to this (holonomic) co-ordinate system yields no physical entities. This resolution is not invariant, either,

* The tensor $g_{\lambda\mu}$ has 21 components, g_{ik} and $r_{ik}^{(s)}$ have together 18, considering that the $r_{ik}^{(s)}$ satisfy the first set of Maxwell equations. The balance is in order, as the three components $g_{(s)(t)}$ are normalization constants and do not count as dynamic variables.

because the parameter net $w^{(s)}$ on the sheets is still arbitrary. The transition between two permissible systems of co-ordinates is identical with a *gauge transformation*.

The impossibility of making the co-ordinate vectors $\partial \xi^\lambda / \partial x^i$ coincide with the space-time world* is a consequence of the relative twisting of the sheets, i.e. of the embedding fields $r_{\lambda\mu}^{(s)}$. The obliqueness of the co-ordinate system (the inclination of the tangential vectors $\partial \xi^\lambda / \partial x^i$ to the normal space n_i^λ , figure 1 a) is thus a measure of the strength of the fields. This slope is given also by the deviation of the path vectors $r_\lambda^{(s)}$ from the gradients of the spaces $w^{(s)} = \text{const.}$ (figure 1 b). The difference

$$h_\lambda^{(s)} = r_\lambda^{(s)} - \partial_\lambda w^{(s)}$$

are the *potentials*. Indeed

$$\partial_{[\lambda} h_{\mu]}^{(s)} = \partial_{[\lambda} r_{\mu]}^{(s)} = r_{\lambda\mu}^{(s)}.$$

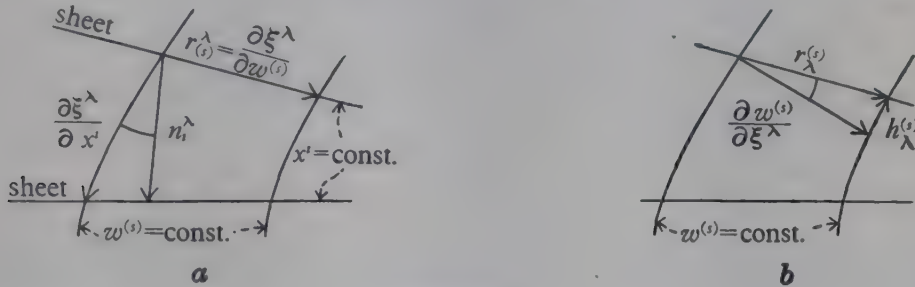


FIGURE 1

But the obliqueness is not an invariant measure of the fields, because it is still dependent on the choice of the net $w^{(s)}$ (the gauge). With a transformation of the $w^{(s)}$ (a gauge transformation) the $h_\lambda^{(s)}$ are changed by gradients. Only when the sheets are parallel ($r_{\lambda\mu}^{(s)} = 0$) can we choose the spaces $w^{(s)} = \text{const.}$ such that they intersect the sheets normally ($h_\lambda^{(s)} = 0$).

Another measure of the obliqueness is given by the mixed components $g_{i(s)}$ of the metric tensor in the *holonomic* co-ordinate system. In the Kaluza-Klein theory the potentials appear in this form.

EQUATIONS OF MOTION. FIELD EQUATIONS

8. SYSTEM OF REFERENCE

(a) We saw that the space normal to the sheets is the space-time world. Of a physical entity, represented by a vector v^λ , we only experience the projection $'v^\lambda$ on to the normal space, the *world component*. We have to project also the six-dimensional laws of motion on to the normal space, in order to predict the fate of this component. It is therefore necessary to know what happens to the world component, if a vector is displaced parallel in the six-dimensional space (induced connexion).

However, the projection of the vector on to the sheet, the *sheet component* $''v^\lambda$, has a well-defined physical meaning, too. As the equations of motion are formulated for the vector $v^\lambda = 'v^\lambda + ''v^\lambda$, the sheet component $''v^\lambda$ appears in the equations of motion for the world component $'v^\lambda$. We shall see that the sheets carry the 'geometry of charge and mass'.

* This property was thought of when referring to the anholonomy of the normal space in § 3 (d).

(b) Therefore we resolve a contravariant vector with respect to the directions of a sexuple $n_i^\lambda, r_{(s)}^\lambda$ ($i = 1, \dots, 4; s = 1, 2$). The duple $r_{(s)}^\lambda$ lies in the sheet, the n_i^λ may generate the normal vector space (the space-time world):

$$g_{\lambda\mu} n_i^\lambda r_{(s)}^\mu = 0.$$

The resolution

$$v^\lambda = n_i^\lambda v^i + r_{(s)}^\lambda v^{(s)}$$

then splits v^λ into world and sheet components. To express in v^λ the components $v^i, v^{(s)}$ with respect to the sexuple, we introduce the reciprocal sexuple of the covariant vectors $n_\lambda^k, r_{(s)}^{(l)}$. The latter is defined so that each of its vectors is orthogonal to five vectors of the first (contravariant) sexuple,

$$\begin{aligned} n_\lambda^k r_{(s)}^\lambda &= 0, & r_{(s)}^{(l)} n_i^\lambda &= 0, \\ n_\lambda^k n_i^\lambda &= \delta_i^k, & r_{(s)}^{(l)} r_{(s)}^\lambda &= \delta_{(s)}^{(l)}. \end{aligned}$$

Then we find $v^i, v^{(s)}$ by projecting v^λ on to the reciprocal sexuple:

$$\begin{aligned} v^i &= n_\lambda^i v^\lambda, \\ v^{(s)} &= r_{(s)}^{(s)} v^\lambda. \end{aligned}$$

For the resolution of a covariant vector the two sets interchange their role:

$$\begin{aligned} v_\lambda &= n_\lambda^i v_i + r_{(s)}^{(s)} v_{(s)}, \\ v_i &= n_i^\lambda v_\lambda, \\ v_{(s)} &= r_{(s)}^\lambda v_\lambda. \end{aligned}$$

(c) The resolution of affinors with respect to the sexuple gives rise to mixed components. The *metric tensor* is an exception. On account of the orthogonality of n_i^λ and $r_{(s)}^\lambda$

$$g_{i(s)} = n_i^\lambda r_{(s)}^\mu g_{\lambda\mu} = 0,$$

and therefore simply

$$g_{\lambda\mu} = n_\lambda^i n_\mu^k g_{ik} + r_{(s)}^{(s)} r_{(t)}^{(t)} g_{(s)(t)}.$$

The sheet component

$$g_{(s)(t)} = r_{(s)}^\lambda r_{(t)}^\mu g_{\lambda\mu}$$

represents the lengths and the direction cosine of the path vectors $r_{(s)}^\lambda$:

$$g_{(1)(1)} = -g_{(2)(2)} = a^2, \quad g_{(1)(2)} = 0.$$

9. INDUCED CONNEXION. FIRST SET OF FIELD EQUATIONS

(a) To find the *covariant derivative*, induced in the normal space, of a vector v_μ , we consider its world component

$$'v_\mu = n_\mu^l v_l \quad (l = 1, \dots, 4)$$

and the covariant derivative of this component in a world direction n_i^λ

$$n_i^\lambda \hat{\nabla}_\lambda 'v_\mu = \hat{\nabla}_i 'v_\mu.$$

This variation does not in general completely belong to the normal space. But the sheet components of the derivative will escape the four-dimensional observer; seen by him, the vector is displaced parallel if the world components of the derivative

$$n_k^\mu \hat{\nabla}_i 'v_\mu = \hat{\nabla}_i 'v_k$$

vanish. This quantity, which lies entirely in the normal space, is therefore defined* as the covariant derivative $\nabla_i v_k$ induced in the space-time world,

$$\nabla_i v_k = \hat{\nabla}_i' v_k = n_i^\lambda n_k^\mu \hat{\nabla}_\lambda (n_\mu^l v_l).$$

Now let us differentiate covariantly the equation of resolution

$$v_\mu = n_\mu^l v_l + r_\mu^{(t)} v_{(t)}$$

and project it on to the normal space. Then the first term on the right-hand side gives just the induced derivative, and we obtain a relation between the derivatives $\hat{\nabla}_i$ and ∇_i :

$$\hat{\nabla}_i v_k = \nabla_i v_k + r_{ik}^{(t)} v_{(t)}. \quad (5)$$

Therefore, if a vector is displaced parallel ($\hat{\nabla}_i v_k = 0$), then, as seen from the space-time world, it appears turned—as we have already made evident (§ 5).

(b) Similarly, we define a connexion, induced in the sheet. It is convenient to extend the definition to *mixed components*, by putting†

$$\nabla_i v_{(s)} = \partial_i v_{(s)},$$

$$\nabla_{(s)} v_i = \partial_{(s)} v_i.$$

Then, as a generalization of (5),

$$\left. \begin{aligned} \hat{\nabla}_a v_b &= \nabla_a v_b - H_{ab}^c v_c, \\ \hat{\nabla}_a v^c &= \nabla_a v^c + H_{ab}^c v^b, \end{aligned} \right\} \quad (6a)$$

where a, b, c can now have each of the six values $i, (s)$, and the summations must be taken over all six values. The affinator H is defined by

$$-H_{ik(s)} = H_{i(s)k} = H_{(s)ik} = r_{ik(s)}, \quad (6b)$$

all other components vanishing. The quantity H_{ab}^c we shall call the *field of constraint*. The equations (6) contain everything we need for the following.

(c) The connexion which we have chosen in the six-dimensional space was defined as 'Riemannian with translation'. The equations characterizing it (with u an arbitrary scalar)

$$\left. \begin{aligned} \hat{\nabla}_\lambda g_{\mu\nu} &= 0 \\ \hat{\nabla}_{[\lambda} \hat{\nabla}_{\mu]} u &= 0 \end{aligned} \right\} \text{Riemannian connexion,}$$

$$\hat{\nabla}_{[\lambda} r_{\mu\nu]}^{(t)} = 0 \quad \text{Killing's equation}$$

can now be projected on to the space-time world with the help of (6). That yields a *first set of field equations* ($i, k, l = 1, \dots, 4; t = 1, 2$)

$$\nabla_i g_{kl} = 0,$$

$$\nabla_{[i} \nabla_{k]} u = H_{ik}^{(t)} \nabla_{(t)} u,$$

$$\nabla_{[i} r_{kl]}^{(t)} = \nabla_{(t)} H_{kl}^{(t)} = 0.$$

The first equation tells us that the induced connexion is 'metric'. However, it is not Riemannian, as the second equation shows. This result will be discussed in § 18. In the last equation we recognize the first quadruple of the Maxwell equations.

* In § 18 (b) another derivation—perhaps more direct—will be given.

† The reason for this choice will be given later on (§ 18 (d)), but our results are independent of these definitions.

10. EQUATIONS OF MOTION. CONSERVATION LAWS

(a) *Equations of motion.* Let the six-dimensional motion of a material point be geodesic. Then, if u^λ represents the six-velocity,

$$u^\mu \hat{\nabla}_\mu u_\lambda = 0,$$

or, referred to the sexuple (a and $b = i, (s)$),

$$u^b \hat{\nabla}_b u_a = 0. \quad (7a)$$

Going over to the induced derivative, using (6a),

$$u^b \nabla_b u_a - H_{ba}^c u^b u_c = 0. \quad (7b)$$

These are six equations. In addition we have the conditions of gauge invariance

$$\partial_{(s)} u_a = 0.$$

The first four equations (7) are ($a = i = 1, \dots, 4$)

$$u^k \nabla_k u_i + 2H_{ik}^{(s)} u^k u_{(s)} = 0.$$

This is the customary form of the equations of motion in an electromagnetic field. Only the last term is the sum ($s = 1, 2$) of two Lorentz forces. These will receive their familiar form, if we put*

$$2H_{ik}^{(s)} = -\frac{1}{e_0} (F_{ik}, G_{ik}),$$

$$u_{(s)} = e_0 \left(\frac{e}{m}, \frac{f}{m} \right),$$

introducing a proportionality factor e_0 with the dimension of a charge. Apart from this factor, we thus have to identify $H_{ik}^{(1)}$ with the electromagnetic field and $u_{(1)}$ with e/m .

The last two equations (7) are ($a = (s) = (1), (2)$)

$$u^k \nabla_k u_{(s)} = u^k \partial_k u_{(s)} = 0$$

and state that the two coupling coefficients e/m and f/m are constants of motion. The existence of integrals of the first order

$$u_{(s)} = \text{const.}$$

of the differential equation of the geodesic has already been quoted as a typical property of spaces which admit of a translation. This property is thus physically a conservation law. Geometrically it means that the six-dimensional world line of a point intersects all sheets at the same angle. Particles of different charges are distinguished by their world lines running at different angles to the sheet, or in other words, by the velocity components parallel to the sheets.

(b) *Conservation laws.* We require that the divergence of a six-dimensional energy-momentum tensor vanishes,

$$\hat{\nabla}_b T_a^b = 0, \quad (8)$$

and that it is gauge invariant,

$$\partial_{(s)} T_a^b = 0.$$

* In the following m stands for mc^2 . Co-ordinates are chosen such that $g_{44} = -c^2$.

The first four equations (8) yield the conservation laws of momentum and energy

$$\nabla_k T_i^k + 2H_{ik}^{(s)} T_{(s)}^k = 0,$$

where thus $T_{(s)}^k$ plays the role of the electric current s^k ,

$$T_{(s)}^k = -e_0(s^k, t^k).$$

The vector t^k stands in the same relation to the field $G_{ik} \sim H_{ik}^{(2)}$ as the current s^k to the electromagnetic field $F_{ik} \sim H_{ik}^{(1)}$.

The last two equations (8) ($a = (s) = (1), (2)$)

$$\nabla_k T_{(s)}^k = 0$$

are the conservation laws for the two currents s^k, t^k . Both mechanical and electrical conservation arise from the one equation (8).

(c) *Identifications.* On account of the equations of motion we could interpret the sheet components of a six-velocity and an energy-momentum tensor. We obtained the identifications

$$\left. \begin{aligned} 2H_{ik}^{(s)} &= -\frac{1}{e_0} (F_{ik}, G_{ik}), \\ u_{(s)} &= e_0 \left(\frac{e}{m}, \frac{f}{m} \right), \\ T_{(s)}^k &= -e_0(s^k, t^k). \end{aligned} \right\} \quad (9)$$

The last two are in harmony with each other. This can be seen, for incoherent matter, by putting

$$T_a^b = -\mu u_a u^b.$$

One then gets, besides

$$T_i^k = -\mu u_i u^k,$$

the expected relations

$$s^k = \frac{e}{m} \mu u^k,$$

$$t^k = \frac{f}{m} \mu u^k.$$

The proportionality factor e_0 depends, as we shall see presently, on the choice of the length scale on the sheets.

11. FIELD EQUATIONS

Let the field Lagrangian be $\frac{1}{2}\dot{R}$, where $\dot{R}$ is the six-dimensional curvature scalar. Variation of the $g_{\lambda\mu}$ gives rise to the field equations

$$\dot{K}_{\lambda}{}^{\mu} = -\kappa T_{\lambda}{}^{\mu}$$

with* $\dot{K}_{\lambda}{}^{\mu} = \frac{1}{2}\dot{R}_{\lambda}{}^{\mu} - \frac{1}{4}\delta_{\lambda}^{\mu}\dot{R}$ and $\kappa = \frac{4\pi}{c^4} \times \text{grav. constant.}$

We have to express the curvature quantities $\dot{R}, \dot{K}$, in terms of those, R, K , of a four-dimensional subspace, the space-time world. (The curvature of the sheets does not

* In the customary notation $K_{\lambda}{}^{\mu} = \frac{1}{2}G_{\lambda}{}^{\mu}$. We had to reserve the letter G for one of the fields of constraint.

enter, because they are pseudo-Euclidean.) This is easy, as we know the relation between the connexion quantities from (6a),

$$\hat{\Gamma}_{ab}^c = \Gamma_{ab}^c + H_{ab}^c.$$

The calculation is straightforward:*

$$\left. \begin{aligned} \hat{K}_i^k &= K_i^k + H_{im}^{(s)} H^{mk}_{(s)} - \frac{1}{4} \delta_i^k H_{nm}^{(s)} H^{mn}_{(s)} = -\kappa T_i^k, \\ \hat{K}_{(s)}^i &= -\frac{1}{2} \nabla_k H^{ik}_{(s)} = -\kappa T_{(s)}^i. \end{aligned} \right\} \quad (10)$$

Indeed, the additional term in the first line is the sum ($s = 1, 2$) of two binomials, both of which have the form of an electromagnetic energy-momentum tensor; the last line is in fact the second Maxwell set. We obtain numerical agreement if we put

$$g_{(1)(1)} = -g_{(2)(2)} = 4\kappa e_0^2. \quad (11)$$

With the identifications (9), derived from the equations of motion, $\hat{K}_{(s)}^i$ gives the equations for the constraining fields (F_{ik}, G_{ik}) and $\hat{K}_i^k$ those of gravitation (K_i^k):

$$\nabla_k F^{ik} = s^i,$$

$$\nabla_k G^{ik} = -t^i,$$

$$K_i^k + \kappa (S_i^{(1)k} - S_i^{(2)k}) = -\kappa T_i^k,$$

with

$$S_i^{(1)k} = F_{im} F^{mk} - \frac{1}{4} \delta_i^k F_{nm} F^{mn},$$

$$S_i^{(2)k} = G_{im} G^{mk} - \frac{1}{4} \delta_i^k G_{nm} G^{mn}.$$

Current and energy-momentum tensor of the two fields $H_{ik}^{(s)} \sim F_{ik}, G_{ik}$ occur in the equations with different signs.† The field F_{ik} is behaving completely Maxwellian. On G_{ik} , the difference in sign has two consequences:

(i) f -charges (sources of G_{ik}) of the same signs attract each other, those with opposite signs repel each other;

(ii) the energy $-S_4^{(2)}$ of the field G_{ik} is negative.

12. FIELD ENERGY. CONFORMAL INVARIANCE

(a) The decomposition of the field $H_{ik}^{(s)}$ into the two components $H_{ik}^{(1)} \sim F_{ik}$ and $H_{ik}^{(2)} \sim G_{ik}$ is actually only a formal procedure. None of the two fields is observable separately. A physical meaning can only be attached to the total field $H_{ik}^{(s)}$.

If, as we found, the two fields F_{ik}, G_{ik} have the same (infinite) range, they result in a single field of the same property. The resulting field is electromagnetic if the first component ($H_{ik}^{(1)} \sim F_{ik}$) is predominant. Indeed, the sources e and f of the fields F_{ik} and G_{ik} are the components of a vector $u_{(s)}$ on the sheet. We can choose the frame of reference (the path vectors) in such a way that $u_{(2)} \sim f$ vanishes (if $u_{(s)}$ is space-like,

* We again refer all quantities to the sexuple $n_i^\lambda, r_{(s)}^\lambda$. The $g_{(s)(t)}$ are not varied, as they are normalization constants.

† This is a consequence of the indefinite metric of the sheets.

$e^2 > f^2$). And with it the field $H_{ik}^{(2)} \sim G_{ik}$ disappears also. The embedding in the six-dimensional space only results in *one* constraining field, the electromagnetic one. When we try, by choosing $e^2 = f^2$, to cancel the field singularities, we extinguish the total field energy.

This result, however, is a consequence of the simplification on which this presentation is based. The isotropy is forced upon the sheets by the non-invariant formulation of the structure axiom (§ 3 (e)). If this is put right, the sheets will cease to be geodesic. They have two distinguished directions, corresponding to the two fields of different range. Then it is no more possible to cancel one of the fields by a particular choice of the frame of reference on the sheet. If we give the field G_{ik} —of negative energy—finite range, we experience, at some distance from the electron, only the electromagnetic field F_{ik} .

Then it is possible to make the field energy of a point source finite by choosing $e^2 = f^2$, without cancelling the whole field energy.

(b) We can replace the *ad hoc* supposition, $e^2 = f^2$, by a more general principle if we postulate that the (six-dimensional) path of a material point is a *null geodesic*. For this is equivalent (according to (9) and (11)) to

$$4\kappa m^2 = e^2 - f^2. \quad (12)$$

Really, f^2 is then very close to e^2 ; for the electron $\sqrt{(\kappa)} m \approx 10^{-20}e$.

Due to the minute difference, the singularities of the field energies of F_{ik} and G_{ik} do not cancel completely. According to sign and order of magnitude, the remainder may possibly make the gravitational energy finite.

In any case the above relation between mass and charges deprives us of the liberty to choose the sources of gravity and of the constraining fields independently. This is perhaps an important step in the direction of the unification of fields.

We get a little encouragement for our null geodesic hypothesis, if we once more fall back on quantum theory. It looks as if a six-dimensional Dirac equation does not admit a mass term;* and that means really the same.—As the strongest argument, however, we consider the fact that the null geodesic assumption makes the theory *conformally invariant*. A length scale does not exist in empty (structureless) space. The material sources, which impose a structure on the space, also gauge the metric tensor.—However, we have not yet formulated the theory in a conformally invariant manner; we supposed space to be Riemannian. Therefore we can only suggest the null geodesic motion as a programme.

But we cannot make use either of the relation $e^2 = f^2$ or of $4\kappa m^2 = e^2 - f^2$ as long as the two fields of constraint F_{ik} , G_{ik} have the same (infinite) range. We can free ourselves from this restriction only by introducing quantum theoretical conceptions.

(c) For the case of equal range of the components, it was shown in (a) that the splitting up of the field $H_{ik}^{(s)}$ into a positive and a negative part is not a significant

* The Dirac matrices transform as a skew-symmetric affnor $\alpha^{\lambda\mu}$. The only possibility to build up by means of these an invariant system of linear differential equations, seems by putting

$$\alpha^{\lambda\mu} \partial_\mu \psi = k^\lambda \psi,$$

where the operator k^λ does not contain the α 's. But then the right-hand side, as a consequence of the commutation relations of the matrices, is already fixed: $k_\lambda = i\partial_\lambda$.

procedure. The negative energy of $H_{ik}^{(2)} \sim G_{ik}$, therefore, cannot give rise to any difficulties; the second field can always be transformed away. It is only the total field—the energy of which is positive—that has any physical meaning attached to it.

The reason for this, as was seen above, is that the two fields $H_{ik}^{(1)}, H_{ik}^{(2)}$ can be transformed into one another by means of a co-ordinate transformation (rotation in the sheet). The two fields are the components of *one* (six-dimensional) field, and it is only the latter which has an invariant meaning.

This feature remains unchanged in the extension of the theory in which the two components are given different ranges. Although two directions in the sheet are now singled out, the covariance of the field equations with respect to rotations in the sheet is not affected.* Still the two components F_{ik}, G_{ik} can be transformed into each other by a rotation in the sheet. One of these fields separately (and its energy) does not represent a physical entity. It can therefore be expected that the difficulties which are usually connected with the concept of a field of negative energy will not appear in this theory. This advantage is due to the fact that, here, the second field is amalgamated with the electromagnetic field.

The situation is very similar to that in quantum electrodynamics, in which a negative energy is attributed to scalar quanta. In both cases it is the time-like components which give a negative contribution to the energy of the field. Here, the negative part of the energy is due to the signature of the second hyperdimension. The null geodesic hypothesis keeps the total energy positive.

EMBEDDING PHYSICS

The intention of this chapter is to clarify geometrically the concept of the 'embedded' space-time world, and to describe more accurately the position of the four-dimensional observer. Potentials and gauge invariance will find their geometric equivalents.

13. ANHOLONOMY

(a) *Ennuples*. In an n -dimensional space—described by the co-ordinates ξ^λ —we introduce a set (ennuple) of n contravariant vectors A_a^λ ($a = 1, \dots, n$). A reciprocal set of covariant vectors A_λ^b ($b = 1, \dots, n$) is associated with it, so that every vector of the latter is orthogonal to $n - 1$ vectors of the first,

$$A_a^\lambda A_\lambda^b = \delta_a^b.$$

We resolve a contravariant vector v^λ with reference to the first set,

$$v^\lambda = A_a^\lambda v^a,$$

and find its components with respect to the ennuple by projecting on to the reciprocal one,

$$v^a = A_\lambda^a v^\lambda.$$

Similarly,

$$v_\lambda = A_\lambda^a v_a,$$

$$v_a = A_a^\lambda v_\lambda.$$

* Compare this with the case of a tensor equation, of which the covariance is not destroyed by the fact that a tensor defines certain distinguished directions.

(b) *Derivatives.* Suppose that for an infinitesimal displacement $\delta\xi^\lambda$ in the direction of a vector A_a^λ ,

$$\delta\xi^\lambda = \delta\alpha \cdot A_a^\lambda,$$

the value of a space function u changes by δu . Then we define a derivative ∂_a in the direction a by

$$\delta u = \delta\alpha \cdot \partial_a u.$$

Derivatives in different directions a, b will in general not be interchangeable. Their commutator will be some linear combination of all derivatives

$$\partial_{[a}\partial_{b]} = \omega_{ab}^c \partial_c.$$

The field ω_{ab}^c is that of *anholonomy*.

(c) *Gap.* The geometric significance of anholonomy is clear. Carrying out in turn two displacements $\delta\alpha \cdot A_a^\lambda$ and $\delta\beta \cdot A_b^\lambda$ in the directions a and b , the result will depend on the order of the displacements. The quadrilateral does not close, the two resultant displacements gape a vector $2\delta\alpha\delta\beta \cdot \partial_{[a}A_{b]}^\lambda$. Therefore we call the vector $\partial_{[a}A_{b]}^\lambda$ the *a,b-gap*. If we resolve it with reference to the ennuple,

$$\partial_{[a}A_{b]}^\lambda = \omega_{ab}^c A_c^\lambda,$$

then the coefficients of anholonomy appear as its components. The quantity

$$\omega_{ab}^c = A_\lambda^c \partial_{[a}A_{b]}^\lambda$$

is thus the c -component of the a, b -gap.

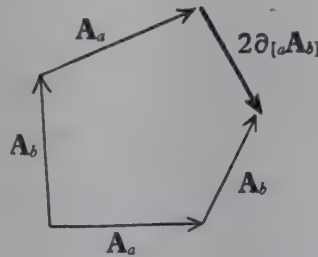


FIGURE 2

(d) *Potentials.* In the reciprocal (covariant) ennuple we come across anholonomy, if we project the rotation $\partial_{[\lambda}A_{\mu]}^c$ of the vector A_μ^c on to the co-ordinate planes,

$$\partial_{[\lambda}A_{\mu]}^c = -A_\lambda^a A_\mu^b \omega_{ab}^c.$$

Now

$$\omega_{ab}^c = -A_a^\lambda A_b^\mu \partial_{[\lambda}A_{\mu]}^c$$

is defined as the (negative) rotation of the vector A_μ^c in the plane a, b . We can therefore call the A_μ^c the potentials of anholonomy.

(e) *Field equations.* Defined thus by

$$\partial_{[a}\partial_{b]} = \omega_{ab}^c \partial_c, \quad (13)$$

ω_{ab}^c is given as c -component of the a, b -gap or as rotation of A_μ^c in the plane a, b ,

$$\begin{aligned} \omega_{ab}^c &= A_\lambda^c \partial_{[a}A_{b]}^\lambda \\ &= -A_a^\lambda A_b^\mu \partial_{[\lambda}A_{\mu]}^c. \end{aligned} \quad (14)$$

As a consequence of definition (13) the anholonomy fulfils the differential equations

$$\partial_{[a}\omega_{bc]}{}^d + 2\omega_{[ab}{}^f\omega_{c]f}{}^d = 0, \tag{15}$$

which are indeed identically satisfied by (14). Apart from the last term these equations look like the first set of the Maxwell equations. It will be shown that they *are* the Maxwell equations.

If all ω disappear we call the ennuple holonomic. Then it is possible to find a system of co-ordinates x^a , so that the local co-ordinate axes coincide with the ennuple: A^λ_a are the tangential vectors $\partial\xi^\lambda/\partial x^a$ of the parameter lines, A^α_λ the gradients $\partial_\lambda x^a$ of the parameter hypersurfaces. The derivative ∂_a in the direction a is then the differentiation with respect to x^a .

(f) *Embedding* ($n = 6$). We divide the vectors of a given sexuple A^λ_a ($a = 1, \dots, 6$) into two groups, $r^\lambda_{(s)}$ ($s = 1, 2$) and n^λ_i ($i = 1, \dots, 4$)—so that each vector of one group is normal to all of the other. This defines a decomposition of the *vector* space: The duple $r^\lambda_{(s)}$ as well as the quadruple n^λ_i generate a vector subspace, which is embedded in the six-dimensional space. (The $r^\lambda_{(s)}$ generate the sheets, the n^λ_i the normal space, the space-time world.)

From the coefficients of the anholonomy we can gather the properties of the embedding. We shall see that in particular the mixed components ($i, k, l = 1, \dots, 4$; $s, t, u = 1, 2$)

$$\omega_{ik}{}^{(s)} \quad \text{and} \quad \omega_{(s)(t)}{}^i$$

characterize the embedding, while the pair

$$\omega_{ik}{}^l \quad \text{and} \quad \omega_{(s)(t)}{}^{(u)},$$

by choosing a suitable frame of reference (i.e. the vectors $n^\lambda_i, r^\lambda_{(s)}$), can be made to vanish.

14. THE SHEETS

(a) What is now the condition that the duple $r^\lambda_{(s)}$ generates a subspace not only of the vector space, but also of the co-ordinate space? In other words, when can we find a family of surfaces (sheets, V_2), so that the duple always touches one of these sheets, that it envelops the sheets? If such a sheet exists, it must be possible to connect two points in it by displacements in the directions of the duple $r^\lambda_{(s)}$. Hence also the gap of the two vectors $r^\lambda_{(1)}, r^\lambda_{(2)}$ must be expressible in terms of $r^\lambda_{(s)}$ only,

$$\partial_{[(s)}r^\lambda_{(t)]} = \omega_{(s)(t)}{}^{(u)}r^\lambda_{(u)},$$

the gap may have no component in a normal direction

$$n^\lambda_i \partial_{[(s)}r^\lambda_{(t)]} = 0.$$

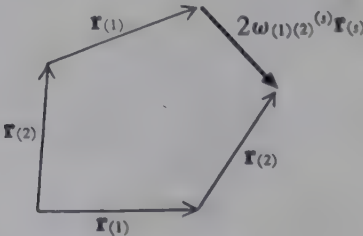


FIGURE 3

That means that the derivatives $\partial_{(s)}$ form a group by themselves (without the ∂_i) and leads to

$$\omega_{(s)(t)}^i = 0. \quad (16a)$$

We then say that the duple is V_2 -generating, or that the subspace of the $r_{(s)}^\lambda$ is holonomic.

(b) Here we deal, according to (3b), with the special case in which (not only the normal components, but) the whole gap vanishes,

$$\partial_{[(s)} r_{(t)}^\lambda = 0.$$

The quadrilateral then closes, and the derivatives $\partial_{(s)}$ form an Abelian group. In this case, besides $\omega_{(s)(t)}^i = 0$, also

$$\omega_{(s)(t)}^{(u)} = 0, \quad (16b)$$

and we can choose a parameter net in every sheet such that the $r_{(s)}^\lambda$ are the respective co-ordinate vectors.

(c) The fact that the duple of the path vectors $r_{(s)}^\lambda$ is V_2 -generating can also be expressed as a property of the normal set, i.e. by the relation

$$\partial_{[\lambda} n_{\mu]}^i = 0.$$

This follows from (14), since, together with $\omega_{(s)(t)}^i$, the sheet components of the rotation of n_λ^i vanish,

$$r_{(s)}^\lambda r_{(t)}^\mu \partial_{[\lambda} n_{\mu]}^i = 0.$$

Then we can always choose the normal quadruple as a set of gradient vectors (§ 15), so that the rotations themselves vanish. This choice of n_λ^i moreover results in

$$\omega_{ik}^i = 0. \quad (16c)$$

15. CO-ORDINATE SYSTEM

It is natural to introduce a co-ordinate system x^i , $w^{(s)}$ in such a way that the sheets—generated by $r_{(s)}^\lambda$ —become co-ordinate surfaces $x^i = \text{const.}$ ($i = 1, \dots, 4$):

$$r_{(s)}^\lambda \partial_\lambda x^i = \partial_{(s)} x^i = 0.$$

This choice of co-ordinates has a physical significance. To each sheet corresponds *one* world point, the co-ordinates of which are just x^i . (Gauge invariant quantities, i.e. those which are constant along a sheet, are therefore functions of the world co-ordinates x^i only.)

The parameter net $w^{(s)}$ on a sheet can be chosen in such a way that the $r_{(s)}^\lambda$ touch the parameter lines:

$$r_{(s)}^\lambda = \frac{\partial \xi^\lambda}{\partial w^{(s)}}.$$

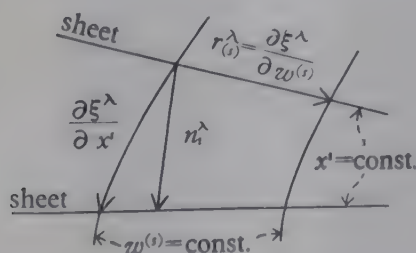


FIGURE 4

The conditions of integrability of these two equations are just those of holonomy (16a, b), $\omega_{(s)}^i = 0$ and $\omega_{(s)}^{(u)} = 0$.

The co-ordinate system is not completely determined by the two conditions. We can

(i) transform the x^i among themselves. This means only a change in the four labels, which are carried by a sheet;

(ii) renumber the net $w^{(s)}$ on the sheets. This causes the spaces $w^{(s)} = \text{const.}$ to intersect the sheets with other angles.

Indeed, we see that the two above equations, which fix the system of co-ordinates, admit only the transformations

$$\begin{aligned} 'x^i &= 'x^i(x) && \text{(co-ordinate transformation),} \\ 'w^{(s)} &= w^{(s)} + \lambda^{(s)}(x) && \text{(gauge transformation).} \end{aligned}$$

We shall soon justify our reference to the second group as a gauge transformation.

The quadruple normal to the sheets can now be chosen as a set of gradient vectors

$$n_\lambda^i = \partial_\lambda x^i. \quad (17)$$

This means that n_i^λ just fits between the sheets, and has as a consequence that the world components ω_{ik}^i of the anholonomy vanish.

16. SPACE-TIME WORLD

(a) *Anholonomy.* The space-time world is carried by the set of vectors n_i^λ normal to the sheets. The normal space is—in contrast to the sheets—not holonomic. The normal quadruple does not define a V_4 in the co-ordinate space. For holonomy would require that $\omega_{ik}^{(s)}$ vanishes, whilst, according to (14),

$$\omega_{ik}^{(s)} = -r_{ik}^{(s)} \quad (18)$$

just defines the field as the anholonomy of the world. The vectors of the normal set do not form quadrilaterals, the field compels them to gape.

However, we could choose the world set in such a way (as gradients) that at least

$$\omega_{ik}^i = 0,$$

i.e. that the i, k -gap

$$\partial_{[i} n_{k]}^\lambda = \omega_{ik}^{(s)} r_{(s)}^\lambda$$

has no world component. The end-points of the two resultant displacements lie in the same sheet, they coincide for the four-dimensional observer.* For him the derivatives ∂_i and ∂_k are commutative, whilst 'in reality'

$$\partial_{[i} \partial_{k]} = \omega_{ik}^{(s)} \partial_{(s)} = -\dot{r}_{ik}^{(s)} \partial_{(s)}.$$

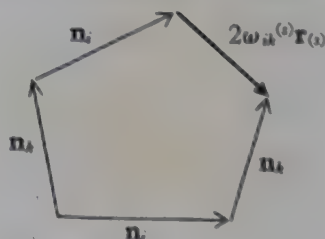


FIGURE 5

* When he encircles a magnetic field, he does not 'in reality' describe a closed path. He experiences a gauge transformation, a translation along the sheet.

The derivatives are really commutative for gauge invariant quantities, namely those which are constant along a sheet.

The field $\omega_{ik}^{(s)} = -r_{ik}^{(s)}$ are now the only non-vanishing components of anholonomy.* The field equations (15) thus become identical with the first half of the Maxwell equations

$$\partial_{[i} r_{kl]}^{(s)} = 0.$$

They appear here as identities, whereas above (§ 3 (d); § 9 (c)) they were produced as a consequence of the structure axiom. The reason is that $r_{\lambda\mu}^{(s)}$ is here defined as $\partial_{[\lambda} r_{\mu]}^{(s)}$ and above as $\hat{\nabla}_{\lambda} r_{\mu}^{(s)}$. The equality of the two quantities is a consequence of the structure axiom.

(b) *World set and co-ordinate axes.* The anholonomy of the space-time world means that there is no co-ordinate system of which four tangential vectors coincide with the world set n_i^λ . The tangential vectors $\partial\xi^\lambda/\partial x^i$ of the co-ordinates defined above certainly do not form the frame of reference of the four-dimensional observer. His world lies normal to the sheet, he goes along the normal from one sheet to the other (from one world point to the neighbouring one). His derivation occurs in the direction of the vectors n_i^λ , which do not touch the spaces $w^{(s)} = \text{const.}$ But with our choice of the world set (which gives the length, indicated in figures 4 and 6a, to n_i^λ) derivation along the normals n_i^λ and differentiation along $w^{(s)} = \text{const.}$ are physically equivalent. Both processes lead to the same sheet, to the same world point—the four-dimensional observer cannot distinguish between them. Yet it is this ‘confusion’ which makes him observe ‘fictitious forces’.

(c) *Potentials.* A measure of the anholonomy, of the field, is thus the inclination of the normals n_i^λ with the tangents $\partial\xi^\lambda/\partial x^i$, the obliqueness of the co-ordinate axes. It can also be measured by the deviation of the path vectors $r_\lambda^{(s)}$ from the gradient vectors $\partial w^{(s)}/\partial\xi^\lambda$, i.e. the two vectors $h_\lambda^{(s)}$. They are just what we previously called potentials of anholonomy, and their components $h_i^{(s)}$ we can now call *embedding potentials*. From (14)

$$\omega_{ik}^{(s)} = -r_{ik}^{(s)} = -\partial_{[i} h_{k]}^{(s)}. \quad (19)$$

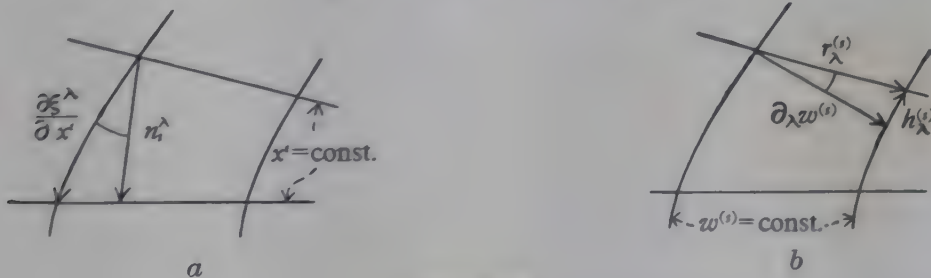


FIGURE 6

The inclination of the spaces $w^{(s)} = \text{const.}$ however, thus also the embedding potentials, depend on the choice of the parameter net $w^{(s)}$. With a transformation

$$w^{(s)} = w^{(s)} + \lambda^{(s)}(x) \quad (20a)$$

$$\text{the potentials transform as } h_i^{(s)} = h_i^{(s)} + \frac{\partial \lambda^{(s)}}{\partial x^i}. \quad (20b)$$

* $\omega_{i(s)}^{(s)}$ vanishes because of the structure axiom. If one refers the definition of the translation $D_{(s)}v_\lambda$ (§ 2 (a)) to the sexuple, one can now show that $D_{(s)} = \partial_{(s)}$.

The fields $r_{ik}^{(s)}$ measure the relative twisting of neighbouring sheets. Only if those are parallel, i.e. $r_{ik}^{(s)} = 0$, can we choose $w^{(s)}$ such that the spaces $w^{(s)} = \text{const.}$ intersect the sheets normally. The potentials $h_i^{(s)}$ will then disappear too.

(d) *Non-gauge-invariant components.* We have now found two rivalling local axes of reference:

- (i) the physical $n_i^\lambda, r_{(s)}^\lambda$, which are not holonomic;
- (ii) the holonomic system of the tangential vectors $\partial \xi^\lambda / \partial x^i, \partial \xi^\lambda / \partial w^{(s)} = r_{(s)}^\lambda$.

With the help of the potentials we can write down the transformation from the holonomic system to the world set. We indicate the components with respect to the holonomic system with a hook $\hook$:

$$\left. \begin{aligned} v_{(s)} &= \check{v}_{(s)}, \\ v_i &= \check{v}_i - h_i^{(t)} \check{v}_{(t)}. \end{aligned} \right\} \tag{21}$$

Especially (because of $\check{\partial}_i = \partial / \partial x^i, \check{\partial}_{(s)} = \partial / \partial w^{(s)}$)

$$\partial_i = \frac{\partial}{\partial x^i} - h_i^{(t)} \frac{\partial}{\partial w^{(t)}}.$$

The world components v_i do not, of course, depend upon the choice of the parameter net on the sheets, but, on the other hand, the holonomic components $\check{v}_i$ do. With a gauge transformation (20 a)

$$'w^{(s)} = w^{(s)} + \lambda^{(s)}(x)$$

they change according to (20 c)

$$' \check{v}_i = \check{v}_i + \frac{\partial \lambda^{(t)}}{\partial x^i} \check{v}_{(t)}.$$

17. PHYSICAL ANALOGIES

The analogy with the behaviour of physical entities is obvious. We have only to show that everything is in order quantitatively.

(a) *Potentials.* Earlier on (§ 10 (c)) we have identified

$$2H_{ik}^{(s)} = -\frac{1}{e_0} (F_{ik}, G_{ik}).$$

Because of

$$H_{ik}^{(s)} = -r_{ik}^{(s)} = -\partial_{[i} h_{k]}^{(s)}$$

we have to put*

$$h_i^{(s)} = \frac{1}{e_0} (A_i, B_i), \tag{22}$$

if we call B_i the potentials of the field G_{ik} .

(b) *Momentum.* If we transform the velocity u_a to the holonomic system,

$$\check{u}_i = u_i + h_i^{(s)} u_{(s)},$$

we obtain,† because of (9), $m\check{u}_i = mu_i + eA_i + fB_i$.

The momenta are the holonomic components of the velocity. (The last term arises from the interaction with the new field.)

* $F_{ik} = \partial_i A_k - \partial_k A_i = 2\partial_{[i} A_{k]}$.

† The factors c are suppressed by our choice of mass unit. Further on we also put $h = \hbar c$ and p instead of pc .

It is hardly necessary to add that a change of the parameter net

$$w^{(s)} = w^{(s)} + \lambda^{(s)}(x)$$

has all the properties of the *gauge transformation*.

(c) *Digression into quantum theory*. Until now we have assumed that all quantities were gauge invariant. The gauge invariance of the fields g_{ik} and $r_{ik}^{(t)}$, their constancy along a sheet,

$$\partial_{(s)} g_{ik} = \partial_{(s)} r_{ik}^{(t)} = 0,$$

was just the content of the structure axiom. It is this limitation which kept us within the framework of a classical theory.

O. Klein taught us how to connect a hyperdimensional theory with a quantum postulate: the points of a sheet are indistinguishable, the position $w^{(s)}$ of a particle on the sheet hence entirely indetermined. Therefore the sheet components $p_{(s)}$ of the particle's momentum is fixed and its wave function is periodical along a sheet,

$$\frac{h}{i} \partial_{(s)} \psi = p_{(s)} \psi.$$

The values of $p_{(s)}$ can be obtained from those of $u_{(s)}$:

$$p_{(s)} = e_0(e, f). \quad (23)$$

The wave function is not gauge invariant. For such quantities, derivation ∂_i along a world direction n_i^λ and differentiation along $w^{(s)} = \text{const.}$ are no longer the same:

$$\begin{aligned} \partial_i &= \frac{\partial}{\partial x^i} - h_i^{(s)} \partial_{(s)} \\ &= \frac{\partial}{\partial x^i} - \frac{i}{\hbar} (eA_i + fB_i). \end{aligned}$$

The difference is just the *charge term* of the Schrödinger equation.

On the basis of the conformally invariant formulation (hypothesis of the null geodesic) we have to write the six-dimensional Schrödinger equation as

$$g^{\lambda\mu} \partial_\lambda \partial_\mu \psi = 0$$

and to identify the second part of the sum

$$g^{ab} \partial_a \partial_b \psi = g^{ik} \partial_i \partial_k \psi + g^{(s)(t)} \partial_{(s)} \partial_{(t)} \psi = 0$$

with the mass term. This leads again to the relation

$$4\kappa m^2 = e^2 - f^2.$$

Charge and mass are therefore differential operators, apart from a proportionality factor, given by

$$e, f \sim \partial_{(s)},$$

$$m^2 \sim \partial^{(s)} \partial_{(s)}.$$

From this point of view the gauge invariance (reality property) of g_{ik} and $r_{ik}^{(t)}$,

$$\partial_{(s)} g_{ik} = \partial_{(s)} r_{ik}^{(t)} = 0,$$

means that gravitons (g_{ik}), photons ($r_{ik}^{(1)}$) and the quanta of the field $r_{ik}^{(2)}$ have neither charge nor mass. The version of the structure axiom which we have chosen for this work could therefore only yield fields of infinite range. To free ourselves from this restriction, we have to leave the boundaries of a purely classical theory. Indeed, the generalization of the structure axiom—mentioned earlier on—does give us the opportunity of making one of the embedding fields complex, i.e. to attribute to it charge, mass and finite range. Only this gives a meaning to the second field. A unitary theory does not seem possible without introducing quantum theoretical concepts. This—classical—treatment therefore can only be incomplete.

18. TORSION OF SPACE-TIME WORLD

(a) *Riemannian connexion* ($\hat{\Gamma}, \hat{\nabla}$). This is defined by two conditions:

(i) A figure may not be deformed when it is displaced parallel. Lengths and angles are preserved if

$$\hat{\nabla}_\lambda g_{\mu\nu} = 0.$$

(ii) The figure as a whole may not be turned. This we check by displacing two vectors parallel along each other: the displaced vectors may not gape. The parallelogram closes if the covariant derivatives of a scalar are commutative,

$$\hat{\nabla}_{[\lambda} \hat{\nabla}_{\mu]} u = 0.$$

Such a connexion is called symmetric. If the indices refer to a holonomic system ξ^λ (hence $\partial_\lambda, \partial_\mu$ commutative), then it follows that $\hat{\Gamma}_{\lambda\mu}^\nu$ is symmetric with respect to λ, μ .

In the latter case the two conditions give

$$\hat{\Gamma}_{\lambda\mu}^\nu = \left\{ \begin{matrix} \nu \\ \lambda\mu \end{matrix} \right\}.$$

We wish to emphasize, however, that symmetry of the connexion and symmetry of $\hat{\Gamma}$ need not be synonymous. If we choose, as a system of reference, an *anholonomic* ennuple A_a^λ with anholonomy ω_{ab}^c , then for a Riemannian connexion

$$\hat{\Gamma}_{ab}^c = \left\{ \begin{matrix} c \\ ab \end{matrix} \right\} + \Omega_{ab}^c,$$

$$\Omega_{abc} = \omega_{abc} - \omega_{acb} - \omega_{bca}.$$

The meaning of the asymmetric additional term is obvious. As the ordinary derivatives ∂_a, ∂_b gape, Ω_{ab}^c takes care that the gap is closed for a parallel displacement. In the case of the set $n_i^\lambda, r_{(s)}^\lambda$ used by us, e.g.

$$\hat{\Gamma}_{ik}^l = \left\{ \begin{matrix} l \\ ik \end{matrix} \right\}, \quad \hat{\Gamma}_{ik}^{(s)} = -r_{ik}^{(s)}.$$

Conversely, the induced connexion is asymmetric, non-Riemannian, but has Γ symmetric.

(b) *Induced connexion* (Γ, ∇). The induction of the connexion is no problem in a co-ordinate subspace (*holonomic* subspace). If we introduce a set of co-ordinates in

the subspace, the components of $g_{\lambda\mu}$ with respect to this co-ordinate system define an 'induced metric tensor'. The Christoffel symbols built up with the latter give the induced connexion.

Also in an *anholonomic* subspace the ennuple of reference defines an induced metric tensor. But we cannot be satisfied with building Christoffel symbols from it. We have to add an Ω , which takes account of the anholonomy.

If, however, the ennuple—like our quadruple n_i^λ —is chosen to be 'almost holonomic', namely so that the gap of n_i^λ and n_k^λ falls outside the subspace ($\omega_{ik}^l = 0$), then, for the embedded observer, his frame of reference is holonomic. He may build up his connexion quantities only from the g_{ik} ,

$$\Gamma_{ik}^l = \left\{ \begin{matrix} l \\ ik \end{matrix} \right\}, \quad \Gamma_{ik}^{(s)} = 0.$$

Yet the connexion* defined by this is not Riemannian. It is true that

$$\nabla_i g_{kl} = 0,$$

but

$$\nabla_{[i} \nabla_{k]} u = \partial_{[i} \partial_{k]} u = \omega_{ik}^{(s)} \partial_{(s)} u = -r_{ik}^{(s)} \partial_{(s)} u.$$

The induced connexion is asymmetric. A 'parallel gap' exists, but it is hidden from the embedded observer, the two end-points lying in the same sheet.

(c) *Torsion*. The asymmetry of connexion is called the torsion of the space. The idea of tracing back the origin of electromagnetism to an asymmetry of connexion is old enough. But earlier theories tried to get the torsion by making Γ asymmetric. We found the torsion—accompanied by a symmetric Γ —from the embedding as a consequence of anholonomy. An essential feature of our procedure is that the torsion is 'harmless'. The observer only experiences a gauge transformation when he encircles field lines. Another advantage is that the torsion satisfies one set of the Maxwell equations identically.

It is, of course, only a difference in terminology whether we interpret the fields $r_{ik}^{(s)}$ as 'twisting of the sheets' (§ 5), as 'anholonomy of the space-time world' (§ 16 (a)) or as its 'torsion'.

(d) *Generalized induced connexion*. It is natural to generalize the expression

$$\Gamma_{ik}^l = \left\{ \begin{matrix} l \\ ik \end{matrix} \right\}$$

for the connexion, induced in the space-time world, to

$$\Gamma_{ab}^c = \left\{ \begin{matrix} c \\ ab \end{matrix} \right\} \quad (a, b, c = i, (s)),$$

and so define also mixed derivatives, as $\nabla_{(s)} v_i$. Comparison with

$$\hat{\Gamma}_{ab}^c = \left\{ \begin{matrix} c \\ ab \end{matrix} \right\} + \Omega_{ab}^c$$

shows that this definition coincides with the one of § 9 (b),

$$\hat{\nabla}_a v_b = \nabla_a v_b - H_{ab}^c v_c,$$

* By comparing the last two pairs of formulae with equation (5) of § 9 (a), one can see that the definition given here coincides with the one there.

if we there choose

$$H_{ab}{}^c = \Omega_{ab}{}^c.$$

Indeed, this definition gives, on account of

$$\Omega_{abc} = \omega_{abc} - \omega_{acb} - \omega_{bca},$$

to the affiner H just the properties, postulated in (6b), because all ω vanish, except for $\omega_{ik}{}^{(s)} = -r_{ik}{}^{(s)}$.

I was—thanks to a Turner and Newall Fellowship—in the fortunate position to finish this publication in Professor L. Rosenfeld's department. I am most grateful for his encouragement and his friendly support. Professor Rosenfeld will find that I have made extensive use of his stimulating remarks. With Professor A. Papapetrou I had interesting discussions. I am much indebted to Professor R. E. Peierls, whose criticism helped me to improve the paper in several places.

I wish to thank Professor H. A. Kramers who made it possible for me to start this work as a Fellow of the Lorentz-fonds.

REFERENCES

- Hoffmann, B. 1947 *Phys. Rev.* **72**, 458.
 Hoffmann, B. 1948 *Phys. Rev.* **73**, 30, 1042.
 Kaluza, Th. 1921 *S.B. preuss. Akad. Wiss.* **54**, 966.
 Klein, O. 1926 *Z. Phys.* **37**, 895.
 Klein, O. 1927 *Z. Phys.* **46**, 188.
 Schouten, J. A. 1935 *Ann. Inst. Poincaré*, **5**, 51.
 Schouten, J. A. & Haantjes, J. 1935 *Ann. Scu. norm. sup. Pisa*, (2), **6**, 175.
 Schouten, J. A. & Struik, D. J. 1938 *Einführung in die neueren Methoden der Differential-geometrie*. Groningen: P. Noordhoff.

One-dimensional dislocations

IV. Dynamics

BY F. C. FRANK AND J. H. VAN DER MERWE

H. H. Wills Physical Laboratory, University of Bristol

(Communicated by N. F. Mott, F.R.S.—Received 7 November 1949)

The equations of motion of the one-dimensional dislocation model are studied, and all solutions representing disturbances propagated without change of form are obtained. These comprise: (i) dislocations, and regular sequences of the same, either of like or of alternating sign, travelling with velocity less than c , the characteristic wave velocity of the system; (ii) anti-dislocations, and sequences of the same, travelling with velocity greater than c ; and (iii) waves of infinitesimal amplitude, belonging to two branches travelling respectively with velocities less than and greater than c . Only dislocations or sequences of dislocations of like sign, and waves of velocity less than c , correspond to stable equilibria. The dislocations exhibit 'relativistic' behaviour. The relevance of anti-dislocations to very fast slip in solids is considered, and rejected.

1. INTRODUCTION

In parts I, II and III of this series (Frank & van der Merwe 1949, 1950) we have dealt with various static properties of a one-dimensional model of a dislocation. In this paper we derive its elementary dynamical properties. We shall obtain only those solutions of our equation which correspond to states of steady motion, but these solutions are so similar to those of the relativistic equations of motion of particles and solid bodies that we feel justified in assuming that this similarity extends to more complicated situations, such as acceleration of dislocations under the influence of forces and collisions between them, and that the behaviour in these cases can be inferred at least qualitatively from the particle results. We believe that this one-dimensional model provides a useful first step in the study of the line dislocations by means of which one now accounts for slip in three-dimensional crystals; it has the additional advantage that with it mechanical models can easily be constructed.

The most important result of this paper is the 'relativistic' relation between the speed and energy of a moving dislocation (equation (21)). A similar result was obtained earlier by Frenkel & Kontorowa (1938), published in the first part of a journal that was never completed and not available in most libraries in this country. This was not known to us until our own investigation was completed. Our own solution is more general than theirs, and we obtain other results not to be found in their paper.

The extension to dislocations in three-dimensional systems has been indicated by one of us in a paper already published (Frank 1949), and has been further developed by Eshelby (1949).

2. THE EQUATION OF MOTION

The model investigated here is a chain of particles of mass m , connected by springs of modulus μ , lying on a substrate with respect to which the potential energy of each particle is a sinusoidal function of position with period a and amplitude $\frac{1}{2}W$. Then, if

$a\zeta_n$ is the displacement of the n th particle from the n th potential trough, n being measured in all cases from an arbitrary zero, the Lagrangian function for a chain containing N particles is

$$L_N = \sum_{n=0}^{N-1} \left\{ \frac{1}{2} m a^2 \dot{\zeta}_n^2 - \frac{1}{2} \mu a^2 (\zeta_{n+1} - \zeta_n - 1/P_0)^2 - \frac{1}{2} W (1 - \cos 2\pi \zeta_n) \right\}, \quad (1)$$

where $1/P_0 = (b-a)/a$ is the 'misfit' between the natural periods of chain and substrate. The n th Lagrangian equation for the system is

$$d(\partial L_N / \partial \dot{\zeta}_n) / dt - \partial L_N / \partial \zeta_n = 0,$$

$$\text{i.e.} \quad m a^2 \partial^2 \zeta_n / \partial t^2 - \mu a^2 (\zeta_{n-1} - 2\zeta_n + \zeta_{n+1}) + \pi W \sin 2\pi \zeta_n = 0. \quad (2)$$

This differs from equation (2) of part I by the addition of an inertia term. If the displacements vary slowly from particle to particle, we may replace the second difference in (2) by a second derivative, obtaining

$$(a^2/c^2) \partial^2 \zeta / \partial t^2 - \partial^2 \zeta / \partial n^2 + (\pi/2l_0^2) \sin 2\pi \zeta = 0, \quad (3)$$

$$\text{where} \quad l_0 = (\mu a^2 / 2W)^{1/2} \quad \text{and} \quad c = a(\mu/m)^{1/2}. \quad (4)$$

Here al_0 is the effective length of a dislocation at rest, as discussed in part I, and c is the characteristic wave velocity for the system, as will be seen in § 3 below.

3. PROPAGATIONAL SOLUTIONS

We shall seek all solutions of equation (3) which represent a disturbance propagated without change of form, i.e. those in which $\zeta(n, t)$ is a function only of

$$z = n - vt/a. \quad (5)$$

Substituting (5) in (3) we have

$$d^2 \zeta / dz^2 = (\pi/2l^2) \sin 2\pi \zeta, \quad (6)$$

$$\text{where} \quad l = l_0(1 - v^2/c^2)^{1/2}. \quad (7)$$

Since (6) is an equation of the same form as I (2), we see immediately that a possible state of motion of the system is one containing moving dislocations which have suffered a 'Lorentz contraction' as expressed by (7). However, we wish to obtain all physical solutions of (6), i.e. all solutions with real z and ζ for real values of the physical constants l_0 , c and v . If l_0 and c are fixed, we have the following two cases to consider:

(i) $0 \leq v^2 \leq c^2$; l is then real, and

$$l_0 \geq l \geq 0;$$

(ii) $c^2 \leq v^2 \leq \infty$; l is then imaginary, and equal to l_1 , say, and

$$0 \leq l_1 \leq \infty.$$

The first integral of (6) may be written in the form

$$d\phi/dz = \pm \pi(1 - k^2 \sin^2 \phi)^{1/2} / lk, \quad (8)$$

$$\text{where} \quad \phi = \pi \zeta - \frac{1}{2} \pi, \quad (9)$$

and k is an integration constant. Then our final solution of (6) is, on integrating (8),

$$z - z_0 = \pm (lk/\pi) \int_0^\phi (1 - k^2 \sin^2 \psi)^{-\frac{1}{2}} d\psi. \quad (10)$$

The second integration constant, z_0 , is of trivial significance, determining only the origin of space and time co-ordinates, and must be real. The real solutions can then be classified as follows.

(i) $0 \leq v^2 \leq c^2$; l is then real, and k must be real:

(i, a) $0 \leq k^2 \leq 1$:

$$z - z_0 = \pm (lk/\pi) F(k, \phi) \quad (11)$$

in the standard notation of elliptic integrals.

(i, b) $k^2 = 1$:

$$z - z_0 = \pm (l/\pi) \ln \tan \frac{1}{2} \pi \zeta. \quad (12)$$

(i, c) $1 \leq k^2 \leq \infty$

$$z - z_0 = \pm (l/\pi) F(1/k, \theta), \quad (13)$$

where $\sin \theta = k \sin \phi$.

These represent respectively: (i, a) a sequence of dislocations of the same sign, recurring at intervals of $(2alk/\pi) K(k)$, (i, b) a single dislocation, and (i, c) a sequence of dislocations of alternating sign, separated from each other by a distance of

$$(2al/\pi) K(1/k),$$

all moving with the velocity v , which is less than c . All these solutions correspond to (and include) the static solutions obtained in part I. The equilibrium in (i, c) is evidently unstable. The case $k = 0$ is trivial, representing an infinite dislocation density. The case $k = \infty$, also, gives only a trivial real solution, for which $\zeta = \frac{1}{2}$, i.e. every particle rests at the top of a potential crest. All three cases lose significance when $v \rightarrow c$, i.e. $l \rightarrow 0$. In fact, as this limit is approached the reduction of the difference equation (2) to the differential equation (3) ceases to be a permissible approximation in the neighbourhood of any dislocation. The approximation is good so long as $l \gg 1$ and fair enough so long as $l > 2$.

(ii) $v^2 \geq c^2$: l is imaginary $= il_1$. Then k must be imaginary $= i\kappa$, say. There can be no real solution with real or complex k , since as ϕ tends to zero, the integral in (10) always approaches zero on the real axis.

(ii, a) $0 \leq \kappa^2 \leq \infty$:

$$\begin{aligned} z - z_0 &= \pm \frac{l_1 \kappa}{\pi} \int_0^\phi (1 + \kappa^2 \sin^2 \psi)^{-\frac{1}{2}} d\psi \\ &= \pm (l_1 k_1/\pi) \{F(k_1, \phi_1) - K(k_1, \phi_1)\}, \end{aligned}$$

or

$$z - z_1 = \pm (l_1 k_1/\pi) F(k_1, \phi_1), \quad (14)$$

where

$$k_1 = \kappa/(1 + \kappa^2)^{\frac{1}{2}},$$

$$\phi_1 = \phi + \frac{1}{2} \pi = \pi \zeta$$

and

$$z_1 = z_0 \mp (l_1 k_1/\pi) K(k_1, \phi_1).$$

(ii, b) $\kappa = \infty$, $k_1 = 1$:

$$\begin{aligned} z - z_0 &= \pm (l_1/\pi) \int_0^\phi \operatorname{cosec} \psi d\psi \\ &= \pm (l_1/\pi) \ln \tan \left\{ \frac{1}{2} \pi (\zeta - \frac{1}{2}) \right\}. \end{aligned} \quad (15)$$

By comparison of (15) with (12) we observe that it represents a state of motion which may be called an 'anti-dislocation'. Each particle remains poised on a potential crest (giving ζ a semi-integral value) except in the neighbourhood of the travelling anti-dislocation, which travels at any speed greater than c . The 'spread' of the anti-dislocation increases with the speed. The equilibrium in this case is clearly unstable. Equation (14) represents a travelling sequence of anti-dislocations, all of the same sign. We defer the question of whether these states of motion can ever be realized until § 5.

None of the cases we have considered above corresponds to the motion of a wave, i.e. an oscillating motion without transport of matter. We shall show that this is because, except for zero amplitude, waves do not propagate themselves without change of form.

For this purpose we examine the complex solutions obtained by taking l imaginary with k real, or conversely. In particular, putting $l = il_1$ in equation (12) or $l_1 = -il$ in (15) we obtain equations which may be written:

$$\tan \frac{1}{2}\pi\zeta = \exp \{ \mp i\pi(z-z_0)/l_1 \} \quad (16)$$

and

$$\tan \{ \frac{1}{2}\pi(\zeta - \frac{1}{2}) \} = \exp \{ \pm i\pi(z-z_0)/l \}. \quad (17)$$

These solutions do actually correspond to the propagation of waves along the chain, but only in the limit of infinitesimal amplitude when the tangent can be replaced by its argument. For whereas equation (16) is a solution of (3), a linear combination of the equations with plus and minus sign, such as

$$\tan \frac{1}{2}\pi\zeta = \cos \{ \pi(z-z_0)/l \}, \quad (16')$$

is *not*. The differential equation is non-linear, and its solutions are not superposable. It becomes linear in the limit when $\sin 2\pi\zeta$ is very small. In particular, if $|\zeta|$ is very small, so that $\sin 2\pi\zeta$ in (3) may be replaced by $2\pi\zeta$, or $|\zeta - \frac{1}{2}|$ is small, so that $\sin 2\pi\zeta$ may be replaced by $2\pi(\frac{1}{2} - \zeta)$, we obtain the solutions

$$\zeta = (2/\pi) \cos \{ \pi(z-z_0)/l_1 \}; \quad |\zeta| \ll 1, \quad (18)$$

$$\zeta - \frac{1}{2} = (2/\pi) \cos \{ \pi(z-z_0)/l \}; \quad |\zeta - \frac{1}{2}| \ll 1. \quad (19)$$

Equation (18) represents oscillations of infinitesimal amplitude about the equilibrium position of each particle, being transmitted as waves of any wave-length $\lambda = 2al_1$, between 0 and infinity, with phase velocity $v = c(1 + \lambda^2/4a^2l_0^2)^{\frac{1}{2}}$. The corresponding group velocity is:

$$u = v - \lambda dv/d\lambda = c(1 + \lambda^2/4a^2l_0^2)^{-\frac{1}{2}}.$$

The frequency has any value from $c/2al_0$ upwards. Equation (19) represents the propagation of waves of wave-length between 0 and $2al_0$, frequency between ∞ and 0, phase velocity $c(1 - \lambda^2/4a^2l_0^2)^{\frac{1}{2}}$ and group velocity $c(1 - \lambda^2/4a^2l_0^2)^{-\frac{1}{2}}$. Since in this case the particles oscillate about their positions of unstable equilibrium this case is of little importance. We mention it nevertheless, to draw attention to the fact that the existence of supersonic anti-dislocations is associated with the existence of a type of wave-motion with supersonic group velocity.

These wave solutions are confined to infinitesimal amplitude. Waves of finite amplitude would not be propagated without change of form. They would also interact with each other when superimposed. Moreover, waves even of infinitesimal amplitude will be scattered by dislocations.

In the case that $\sin 2\pi\zeta$ is small, the difference-differential equation (2) need not be approximated to a differential equation (3), but can be solved as such. When $|\zeta|$ is small, so that $\sin 2\pi\zeta$ may be replaced by $2\pi\zeta$ (i.e. the particles oscillate about their positions of equilibrium) (2) becomes the well-known equation of the band-pass filter (see, for example, Kármán & Biot 1940). Waves are transmitted undamped with frequencies from ν_1 to ν_2 where

$$\nu_1 = (W/2ma^2)^{\frac{1}{2}} = c/2al_0,$$

$$\nu_0 = (\mu/\pi^2m)^{\frac{1}{2}} = c/\pi a$$

and

$$\nu_2 = (\nu_1^2 + \nu_0^2)^{\frac{1}{2}}.$$

ν_1 is the frequency at which all particles oscillate in phase and ν_2 the frequency at which alternate atoms vibrate in opposite phase. ν_0 would be the frequency of this latter vibration for the chain on a flat substrate. The phase velocity for a wave-length λ is

$$\begin{aligned} v &= \lambda \{ \nu_0^2 \sin^2 (\pi a / \lambda) + \nu_1^2 \}^{\frac{1}{2}} \\ &= c \{ (\lambda / \pi a)^2 \sin^2 (\pi a / \lambda) + (\lambda / 2al_0)^2 \}^{\frac{1}{2}}. \end{aligned}$$

For long waves the first term in the root tends to unity, as given by the approximate equation (3).

The group velocity is

$$\begin{aligned} u &= \frac{1}{2} \pi a \nu_0^2 \sin (2\pi a / \lambda) \{ \nu_0^2 \sin^2 (\pi a / \lambda) + \nu_1^2 \}^{-\frac{1}{2}} \\ &= c (\lambda / 2\pi a) \sin (2\pi a / \lambda) \{ (\lambda / \pi a)^2 \sin^2 (\pi a / \lambda) + (\lambda / 2al_0)^2 \}^{-\frac{1}{2}}, \end{aligned}$$

which, instead of tending to c for short waves, as in the solution of equation (3), goes to zero when λ becomes $2a$, after passing through a flat maximum value less than c :

$$u_{\max.} = c \{ (1 + \nu_1^2 / \nu_0^2)^{\frac{1}{2}} - \nu_1 / \nu_0 \}.$$

4. ENERGY

Having obtained the propagational solutions of the equation of motion, we proceed to calculate the energies connected with moving dislocations. It was first shown by Frenkel & Kontorowa that the energy of a moving single dislocation exhibits a striking analogy with the energy of a particle in relativistic mechanics. We shall, however, consider the more general case of a dislocation in a moving sequence. The energy of a single dislocation is then obtained as a special case.

The kinetic energy T of a dislocation in a sequence moving with constant velocity is given by the first term in (1), where N is now equal to P , the number of particles contained in each dislocation. If the displacement varies slowly from particle to particle we may replace the summation by an integral. Using (5) and (8), it is then seen that

$$\begin{aligned} T &= \frac{1}{2} m v^2 \int_0^P \left(\frac{d\zeta}{dz} \right)^2 dz \left(= \frac{1}{2} m v^2 \int_0^1 \frac{d\zeta}{dz} d\zeta \right) \\ &= m E(k) v^2 / \pi k l, \end{aligned}$$

where $E(k) = \int_0^{\frac{1}{2}\pi} (1 - k^2 \sin^2 \psi)^{\frac{1}{2}} d\psi$ is the complete elliptic integral of the second kind.

Substituting for m from (4), we get

$$T = 2Wl_0^2 E(k) v^2 / \pi k c^2 l. \quad (20)$$

If the summation is taken over all the particles P contained in a dislocation, then the last two sums in (1) similarly give minus the *potential* energy of the dislocation. When the relative displacements of neighbouring particles vary slowly from particle to particle $\zeta_{n+1} - \zeta_n$ may be replaced by $d\zeta/dn$ and the summation by an integral. From (4) and (5) it follows then that the potential energy,

$$U = Wl_0^2 \int_0^1 \left(\frac{d\zeta}{dz} - \frac{1}{P_0} \right)^2 \frac{dz}{d\zeta} d\zeta + \frac{1}{2} W \int_0^1 (1 - \cos 2\pi\zeta) \frac{dz}{d\zeta} d\zeta.$$

If we substitute for $dz/d\zeta$ and carry out the integration this becomes

$$U = W \left\{ \frac{2E(k)}{\pi k l} l_0^2 \left(2 - \frac{v^2}{c^2} \right) + \frac{2l(k^2 - 1) K(k)}{\pi k} - \frac{2l_0^2}{P_0} + \frac{Pl_0^2}{P_0^2} \right\}. \quad (21)$$

The sum of the potential and kinetic energies then gives the *total* energy of the dislocation,

$$H_1 = W \left\{ \frac{4E(k) l_0^2}{\pi k l} + \frac{2l(k^2 - 1) K(k)}{\pi k} - \frac{2l_0^2}{P_0} + \frac{Pl_0^2}{P_0^2} \right\}. \quad (22)$$

In the case of a single dislocation ($k = 1$) remote from all others, in a chain of particles whose equilibrium spacing is the same as that of the substrate ($1/P_0 = 0$), this becomes

$$H = 4Wl_0^2/\pi l = 4Wl_0/\pi(1 - v^2/c^2)^{\frac{1}{2}}. \quad (23)$$

This is the result exhibiting the striking relation between speed and energy. Under these same conditions we have

$$T = 2Wl_0/\pi(c^2/v^2 - 1)^{\frac{1}{2}}$$

and

$$U = 2Wl_0(2 - v^2/c^2)/\pi(1 - v^2/c^2)^{\frac{1}{2}}.$$

To show that (23) is completely analogous to the corresponding relationship in relativistic mechanics, we call the (potential) energy of a single dislocation at rest its rest energy $E_0 = 4Wl_0/\pi$, and set up the expressions for

$$\text{the rest mass} \quad M_0 = E_0/c^2 = 4Wl_0/\pi c^2,$$

$$\text{the apparent mass} \quad M = M_0(1 - v^2/c^2)^{-\frac{1}{2}} = 4Wl_0^2/\pi c^2 l,$$

$$\text{and the total energy} \quad H = Mc^2 = 4Wl_0^2/\pi l \quad (24)$$

according to relativistic equations. From (23) and (24) it is seen that our statement is correct.

The total energy of a single anti-dislocation in a chain of particles whose equilibrium spacing is equal to the substrate wave-length, can be likewise shown to be

$$H_a = 4Wl_0^2/\pi l_1 = 4Wl_0/\pi(v^2/c^2 - 1)^{\frac{1}{2}}, \quad (25)$$

where the zero of energy is now taken to be the state in which all particles are on potential crests.

5. ANTI-DISLOCATIONS: DISCUSSION

We are now in a position to consider whether anti-dislocations have physical significance. First, we may note that the speed with which the chain moves over the substrate, by the passage of dislocations, must be less than the characteristic wave velocity c . For the velocity of translation is the product of dislocation density and dislocation velocity: the latter must be less than c , and a dislocation density exceeding one per substrate period is not possible. On the other hand, there is no physical impossibility in a translation velocity greater than c . The chain could certainly be made to travel faster than this and then brought into contact with the substrate. It is reasonable to suppose that the particles would then spend more of the time out of the potential troughs than in them, so that the suggestion readily presents itself that anti-dislocations do represent the conceptual elements into which the resulting state of motion can be analyzed. To establish this, we should need to show that the anti-dislocations, at sufficiently high speed and density, are stable against small perturbations. This we have not done: it is quite obvious that at low speed and density they are not. Supposing that at sufficient speed and density the anti-dislocations do persist, the resulting motion is a strange one, since they have a negative effective mass (their energy decreases with increase of velocity). Thus under a force which would be expected to accelerate the slip, the anti-dislocations will slow down, and contract: an effective acceleration of the slip may still occur, since with a large amount of energy stored in each anti-dislocation, it will probably undergo multiplication (pair formation).

The possibility appears to exist that slip commencing by the motion of dislocations, continuing under high stress could develop into supersonic slip by anti-dislocations. Multiplication of dislocations would occur when their energy was high: then, at a sufficiently high density the distinction between a succession of dislocations and a succession of antidislocations becomes obscured. The corresponding energy is infinite, according to our equations, but this is obviously a defect of the approximation, like the infinite 'sonic barrier' in the theory of flight.

The purpose of this speculation is to see whether our model gives any support to a picture of crystal slip which was generally held prior to the development of the theory of dislocations; one in which a general disengagement of the interlocking atomic meshes occurs over a whole crystal plane. We infer that a state of motion approaching this would develop from dislocation slip, but that in its development the atomic motions would be by no means uniform, and would involve very high kinetic energies. Beyond a doubt, any real crystal would then melt in the neighbourhood of the active glide plane, and spread the relative motion over a number of atomic planes so that none was travelling faster than sound relative to its neighbours. This does not appear to correspond to any actual process of plastic deformation of metals under ordinary, relatively slow, methods of test.

One of us (J. H. van der Merwe) has to thank the South African Council for Scientific and Industrial Research for a grant and special leave which rendered it possible to do this research.

REFERENCES

- Frank, F. C. 1949 *Proc. Phys. Soc. A*, **62**, 131.
 Frank, F. C. & van der Merwe, J. H. 1949 *Proc. Roy. Soc. A*, **198**, 205 (part I) and 216 (part II).
 Frank, F. C. & van der Merwe, J. H. 1950 *Proc. Roy. Soc. A*, **200**, 125.
 Frenkel, J. & Kontorowa, T. 1938 *Phys. Z. Sowjet.* **13**, 1.
 Eshelby, J. D. 1949 *Proc. Phys. Soc. A*, **62**, 307.
 Kármán, T. v. & Biot, M. A. 1940 *Mathematical methods in engineering*, p. 463. New York: McGraw-Hill.

Integration of the laminar boundary layer equation.

I. Motion of an elliptic cylinder. Separation

BY D. MEKSYN,* *Department of Mathematics, King's College, London*

(Communicated by G. Temple, F.R.S.—Received 25 October 1949)

The integration of the equation of the boundary layer

$$\frac{\partial^3 f}{\partial \sigma^3} + f \frac{\partial^2 f}{\partial \sigma^2} = \lambda \left[1 - \left(\frac{\partial f}{\partial \sigma} \right)^2 \right] + 2\gamma \left[\frac{\partial f}{\partial \sigma} \frac{\partial^2 f}{\partial \sigma \partial \gamma} - \frac{\partial f}{\partial \gamma} \frac{\partial^2 f}{\partial \sigma^2} \right]$$

is discussed, and applied to Schubauer's measurements of a flow round an elliptic cylinder.

Two methods of step-by-step integration are indicated.

The separation point is found by approaching it from downstream of separation; and in the second approximation its position is at $x = 2.00$ as against Schubauer's value $x = 1.99$.

It is also shown that in order that physical separation (without singularities) may take place, the curve of velocity $u(x)$ outside the boundary layer must have a point of inflexion.

The equation is approximately integrated from the position of minimum pressure, and compared with Hartree's results.

The equation is approximately integrated downstream of separation for a certain distance, and the evaluated velocities are compared with the observed values.

1. INTRODUCTION

In a previous paper (Meksyn 1948) the equations of motion of the laminar boundary layer in two dimensions were reduced to

$$\frac{\partial^3 f}{\partial \sigma^3} + f \frac{\partial^2 f}{\partial \sigma^2} = \lambda \left[1 - \left(\frac{\partial f}{\partial \sigma} \right)^2 \right] + 2\gamma \left[\frac{\partial f}{\partial \sigma} \frac{\partial^2 f}{\partial \sigma \partial \gamma} - \frac{\partial f}{\partial \gamma} \frac{\partial^2 f}{\partial \sigma^2} \right], \quad (1.1)$$

where γ is the velocity potential of the inviscid flow outside the boundary layer for velocity unity at infinity and $\gamma = 0$ at the forward stagnation point,

$$\sigma = \frac{1}{2} \left(\frac{2u_0}{\nu\gamma} \right)^{\frac{1}{2}} \beta, \text{ where}$$

β is the stream function of the external inviscid flow for velocity unity at infinity,

u_0 the velocity at infinity,

* Now at the Department of Mathematics, Imperial College of Science and Technology.

$\psi = (2\nu u_0 \gamma)^{\frac{1}{2}} f(\sigma, \gamma)$ is the stream function of the viscous flow,

$\lambda = -\gamma \left(\frac{\partial \log h^2}{\partial \gamma} \right)_0$, the value of γ being taken on the boundary

h^2 being the square of the velocity outside the boundary layer.

The equation (1.1) has to be integrated under the boundary conditions

$$\left. \begin{aligned} f(\sigma, \gamma) = \frac{\partial f}{\partial \sigma} = 0 & \quad \text{for } \sigma = 0, \\ \frac{\partial f}{\partial \sigma} = 1 & \quad \text{for } \sigma = \infty. \end{aligned} \right\} \quad (1.2)$$

The equation (1.1), when the term proportional to γ is disregarded, was applied to the motion of a circular and elliptic cylinders; for the outside flow it was assumed the usual theoretical expression for inviscid flow; and it was shown that at the front part of the cylinders the theory is in good agreement with experimental results.

These terms, however, could not be disregarded in the general case; and the aim of the present paper is to discuss the integration of the complete equation (1.1) for measured pressure distribution; it is applied to Schubauer's (1935) measurements of flow round an elliptic cylinder.

2. ASYMPTOTIC BEHAVIOUR OF THE INTEGRALS

Differentiating (1.1) with respect to σ we obtain

$$f^{iv} + \left(f + 2\gamma \frac{\partial f}{\partial \gamma} \right) f''' + (1 + 2\lambda) f' f'' - 2\gamma f' \frac{\partial f''}{\partial \gamma} = 0. \quad (2.1)$$

Now, when $\sigma = \infty$, $\frac{\partial f}{\partial \sigma} \sim 1$, $f \sim \sigma + k(\gamma)$, (2.2)

where $k(\gamma)$ is independent of σ but may depend on γ , whence from (2.1) and (2.2)

$$f^{iv} + (\sigma + l(\gamma)) f''' + (1 + 2\lambda) f'' - 2 \frac{\partial f''}{\partial \gamma} \sim 0. \quad (2.3)$$

where $l(\gamma) = k(\gamma) + 2\gamma \frac{\partial k}{\partial \gamma}$.

Putting, as usual, $f'' = C e^{\xi}$, we find for large σ

$$f'' \sim A(\gamma) \sigma^{2\lambda - l^2} \exp \left(-\frac{\sigma^2}{2} - l\sigma \right). \quad (2.4)$$

3. AN INTEGRAL OF THE EQUATION

$$f''' + f f'' = \lambda(1 - f'^2) + 2\gamma \left(f' \frac{\partial f'}{\partial \gamma} - \frac{\partial f}{\partial \gamma} f'' \right). \quad (3.1)$$

It is assumed that the equation has a unique integral $f(\sigma, \gamma)$ satisfying the required boundary conditions. If we substitute this integral $f(\sigma, \gamma)$, assumed known, in the right-hand side of (3.1) and in f , in the term $f f''$, we can consider that the right-hand

side of (3.1), and f in ff'' are known functions, and integrate (3.1) as a linear non-homogeneous equation; whence

$$f''(\sigma, \gamma) = a e^{-F(\sigma, \gamma)} + e^{-F(\sigma, \gamma)} \int_0^\sigma \left[\lambda(1 - f'^2) + 2\gamma \left(f' \frac{\partial f'}{\partial \gamma} - \frac{\partial f}{\partial \gamma} f'' \right) \right] e^{F(\sigma, \gamma)} d\sigma, \quad (3.2)$$

where

$$F(\sigma, \gamma) = \int_0^\sigma f(\sigma, \gamma) d\sigma.$$

It is obvious that

$$a = f''(0, \gamma).$$

The expression (3.2) represents a kind of integral equation for the unknown function $f(\sigma, \gamma)$, the function $a(\gamma)$ being found from the condition

$$f'(\infty, \gamma) = 1.$$

The equation (3.2), under the above condition, can be integrated by several methods depending on the sign of λ .

We first consider a simple, approximate, short-cut solution, which is applicable when $\lambda < 0$, i.e. for retarded flow.

Assume that the function $f(\sigma, \gamma)$ can be expanded in powers of σ , where the coefficients are functions of γ ; let

$$f(\sigma, \gamma) = \sum_{n=2}^{\infty} \frac{a_n \sigma^n}{n!}. \quad (3.3)$$

Substituting (3.3) in (3.1) we easily find

$$a_2 = a, \quad a_3 = \lambda, \quad a_4 = 0, \quad a_5 = -a^2(1 + 2\lambda) + 2\gamma\lambda'aa', \quad (3.4)$$

where primes denote differentiation with respect to γ .

From (3.2), (3.3) and (3.4) we get after few transformations

$$\begin{aligned} f''(\sigma, \gamma) = & \left\{ a + \lambda \left[\sigma - \frac{a^2}{3} \sigma^3 + \frac{a(1-6\lambda)}{4!} \sigma^4 + \frac{\lambda(1-6\lambda)}{5!} \sigma^5 - \right. \right. \\ & - \frac{(1-2\lambda)a^3 + 2\gamma a^2 a'}{72} \sigma^6 \left. \right] + 2\gamma \left[\frac{aa'}{6} \sigma^3 + \frac{a\lambda'}{12} \sigma^4 + \right. \\ & + \frac{\lambda\lambda'}{60} \sigma^5 + \frac{(15-6\lambda)a^2 a' - 2\gamma aa'^2 - 8\lambda'a^3 + 8\gamma a^2 a''}{6!} \sigma^6 + \\ & + \left. \left. \frac{(31-12\lambda)\lambda aa' + [(45-68\lambda)\lambda' + 2\gamma\lambda'']a^2 + 18\gamma\lambda a'^2 + 18\gamma\lambda aa''}{7!} \sigma^7 + \dots \right] \right\} \\ & \times e^{-F(\sigma, \gamma)}, \end{aligned} \quad (3.5)$$

where

$$F(\sigma, \gamma) = \frac{a\sigma^3}{3!} + \frac{\lambda\sigma^4}{4!} + \frac{-(1+2\lambda)a^2 + 2\gamma aa'}{6!} \sigma^6 + \dots \quad (3.6)$$

From the boundary condition $f'(\infty, \gamma) = 1$. (3.7)

and (3.5) we obtain an ordinary differential equation for the unknown function $a(\gamma)$.

The important point to notice in connexion with this integration is that in certain cases the first few terms in (3.5) only are of importance; this is because the expression $e^{-F(\sigma, \gamma)}$ rapidly tends to zero even for small values of σ .

4. SELF-CONSISTENCY OF THE SOLUTION

It is important to show that the expression (3.2) is self-consistent, i.e. that, if the correct form of $f(\sigma, \gamma)$ is substituted in the right-hand side of (3.2), it will lead to an expression of the same form.

The asymptotic expression for $f''(\sigma, \gamma)$ is (2.4), namely,

$$\frac{\partial^2 f}{\partial \sigma^2} \sim A(\gamma) \sigma^{2\lambda-l^2} \exp\left(-\frac{\sigma^2}{2} - l\sigma\right). \quad (4.1)$$

Since for large σ

$$1 - f = \int_{\sigma}^{\infty} A \sigma^{2\lambda-l^2} \exp\left(-\frac{\sigma^2}{2} - l\sigma\right) d\sigma \sim A \sigma^{2\lambda-l^2-1} \exp\left(-\frac{\sigma^2}{2} - l\sigma\right),$$

$$1 + f' \sim 2,$$

it is easy to show that the term in (3.2) proportional to $(1 - f'^2)$ will lead to an expression for f'' similar to (4.1).

It can also be shown that, if (4.1) is substituted in (3.2) in the terms

$$f' \frac{\partial f'}{\partial \gamma} - \frac{\partial f}{\partial \gamma} f'', \quad (4.2)$$

then the terms proportional to $\sigma^{2\lambda-l^2}$ will cancel out, and (4.2) will lead to an expression for f'' which will retain the exponent $2\lambda - l^2$ in σ ; thus the solution is self-consistent.

It is interesting to notice that, if we separated the terms proportional to γ in (3.1), and integrated the equation

$$f''' + \left(f + 2\gamma \frac{\partial f}{\partial \gamma}\right) f'' = \lambda(1 - f'^2) + 2\gamma \frac{\partial f'}{\partial \gamma} f'$$

by the above method, we should find that it leads to inconstant results; namely, the exponent in σ does not retain its value.

ELLIPTIC CYLINDER

5. PRELIMINARY RESULTS

The flow round an elliptic cylinder was measured by Schubauer (1935), the dimensions of the ellipse being

$$2a = 11.78 \text{ in.}, \quad 2b = 3.98 \text{ in.}$$

The direction of the main flow was along the major axis of the ellipse, the velocity at infinity being $u_0 = 11.5 \text{ ft./sec.}$

The co-ordinate x was measured along the boundary starting from the forward stagnation point, and the distance expressed as a multiple of the minor axis was denoted by x ; the co-ordinate y is taken normal to the boundary.

Schubauer measured the pressure on the boundary, and the velocity u in the boundary layer at many points, and found the position of separation.

The values of u , $\partial u/\partial x$, $\partial^2 u/\partial x^2$ were also evaluated.
We have now to express λ and $\partial \lambda/\partial \gamma$ by means of Schubauer's data.
Let $\bar{u}$ be the velocity outside the boundary layer for speed unity at infinity.
Since γ is the velocity potential, the following results hold on the boundary, viz.

$$\left. \begin{aligned} \frac{\partial \bar{u}}{\partial \gamma} &= \frac{\partial \bar{u}}{\partial x} \frac{\partial x}{\partial \gamma} = \frac{1}{\bar{u}} \frac{\partial \bar{u}}{\partial x}, \\ \lambda &= -\gamma \frac{\partial \log h^2}{\partial \gamma} = -\frac{2\gamma}{\bar{u}^2} \frac{\partial \bar{u}}{\gamma x}, \\ \frac{\partial \lambda}{\partial \gamma} &= -\frac{2\bar{u}'}{\bar{u}^2} + \frac{4\phi \bar{u}'^2}{\bar{u}''} - \frac{2\phi \bar{u}''}{\bar{u}^3}, \\ \gamma \frac{\partial \lambda}{\partial \gamma} &= \frac{2\phi}{\bar{u}^2} \left[-\bar{u}'(1+\lambda) - \frac{\lambda \bar{u}''}{\bar{u}} \right], \end{aligned} \right\} \tag{5.1}$$

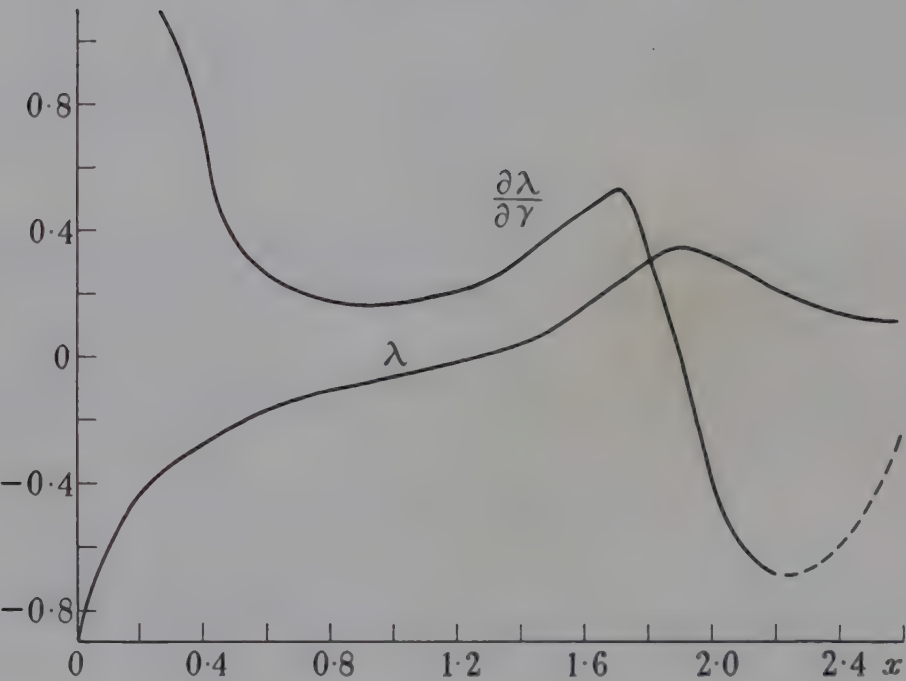


FIGURE 1

TABLE 1

x	γ	λ	$\partial \lambda/\partial \gamma$
0.175	0.080	-0.463	1.29
1.250	1.392	-0.003	0.216
1.350	1.522	0.021	0.269
1.457	1.660	0.066	0.339
1.600	1.844	0.157	0.458
1.700	1.972	0.228	0.529
1.832	2.139	0.326	0.227
1.900	2.224	0.341	-0.041
2.000	2.349	0.308	-0.408
2.100	2.473	0.266	-0.608
2.196	2.591	0.204	-0.687
2.568	3.046	0.106	-0.228
2.937	3.498	0.002	-0.366
3.307	3.951	0.159	-0.381

where primes denote differentiation with respect to x and $\phi \equiv \gamma$ (the notation ϕ is used by Schubauer).

In table 1 are given the values of γ , λ and $\partial\lambda/\partial\gamma$ for several x .

The curves of λ and λ' against x are plotted in figure 1.

6. SEPARATION

The usual procedure for finding the separation point is to integrate the equation of the boundary layer starting from the forward stagnation point, obtaining the values of $a = f''(0, \gamma)$; and separation corresponds to $a = 0$.

This procedure, however, gives rise to very great difficulties, and different methods of integration lead to different results.

The origin of these difficulties is as follows.

From experimental evidence we can conclude that round the point of separation a has roughly the form as in figure 2; the analytical expression of a as a function of x is obviously very complicated. This is connected with the complicated change of λ .

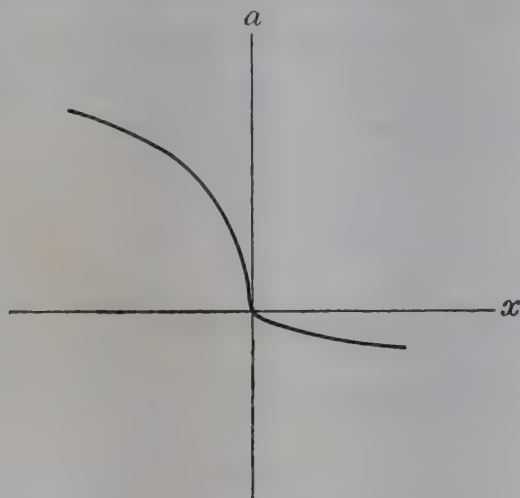


FIGURE 2

The other difficulty is due to the form of the expression (3.5) near separation; this will be considered in some detail.

For $a = 0$, the equations (3.5) and (3.7) become

$$\int_0^\infty \left\{ \lambda\sigma + 2\gamma \left[\frac{\lambda\lambda'}{60} \sigma^5 + \frac{18\lambda\gamma a'^2}{7!} \sigma^7 \right] \right\} e^{-\lambda\sigma^4/24} d\sigma = 1, \quad (6.1)$$

since we assume on physical grounds that the derivatives of a remain finite and that the term $\frac{\lambda^2(1-6\lambda)}{5!} \sigma^5$ may be disregarded as insignificant.

If the terms proportional to γ were disregarded in (3.5), separation would correspond to $\lambda = 0.199$ (Hartree 1937); but since the disregarded terms, when $a \neq 0$, are mainly negative, the point of separation should correspond to a larger value of λ ; that is only possible if $\lambda' < 0$ at the separation point, as can be seen from (6.1).

Since upstream of separation $\lambda' > 0$, we obtain the following result:

If the boundary-layer approximations stay valid, then in order that physical separation (without singularities) may take place it is necessary that the curve $u(x)$ shall have a point of inflexion.

If λ' remains positive, as, for instance, in the case of the usual theoretical solution for inviscid flow, separation can only take place if the disregarded term (3.5) proportional to aa' remains finite and negative right up to and including the separation point, i.e. if the function $f(\sigma, \gamma)$ has a singularity at the separation point.

In the case of Schubauer's measurements $\lambda' = 0$ at $x = 1.900$. If the integration is started from the forward stagnation point, and the obtained values of $a = f''(0, \gamma)$ are smaller than their correct values, separation will take place at $x < 1.900$.

It would, however, require a high degree of approximation for a , particularly when it begins to change rapidly, to give the correct position of separation.

There is fine interplay of values of the principal terms in (3.5), namely,

$$\frac{aa'}{6} \sigma^3 + \frac{a\lambda'}{12} \sigma^4 + \frac{\lambda\lambda'}{60} \sigma^5$$

in the region $x = 1.900$ to 2.000 which makes it so hard to get the right position for separation by the usual method of integration which starts from the forward stagnation point; and, if separation is not obtained up to $x = 2.000$, a will begin to increase, as can be seen from the behaviour of λ and λ' (figure 1).

The position of separation, can, however, be easily obtained if we approach the separation point from downstream of separation.

In order to elucidate the method we shall give a simple example. Consider the equation

$$2 - x + 2a + a'^2 + \dots = 0, \quad (6.2)$$

where a is a function of x .

We have to find the value of x where a vanishes. We can find it by two methods of successive approximations.

(1) Disregarding a'^2 , we obtain that a vanishes at

$$x = 2.$$

Differentiating the equation $2 - x + 2a = 0$

we get that a' is equal to $\frac{1}{2}$, whence in the second approximation a vanishes at

$$x = 2 + a'^2 = 2\frac{1}{4}.$$

(2) As in the former case, in the first approximation

$$x = 2.$$

To find a better approximation consider the point

$$x = 2.1 + \Delta x \quad \text{and} \quad a = b\Delta x. \quad (6.3)$$

Substituting (6.3) in (6.2) and equating to zero the independent terms, and the terms proportional to Δx , we find

$$b = \frac{1}{2}, \quad a' = 0.31,$$

which do not agree.

Consider next the point $x = 2.3 + \Delta x, \quad a = b\Delta x.$

By the same procedure $b = \frac{1}{2}, \quad a' = 0.548.$

Plotting now the obtained values of b and a' we can find the point where they are equal; which is the required solution.

In what follows use was made of the second procedure.

To evaluate the position of separation use is made of the equations (6.1) and (3.5), (3.7).

In the *first approximation* we disregard all derivatives of a . From (6.1), integrating in gamma functions, and making use of the curves in figure 1, we easily find that separation is at

$$x \approx 1.900.$$

In the *second approximation* we retain only the first derivative of a , i.e. a' .

The expression (3.7) is a differential equation in a which can be easily integrated close and downstream of separation.

The term $\exp(-a\sigma^3/6)$ is expanded in powers of σ and in the final expression terms up to σ^7 are retained.

Assuming that separation takes place at γ_0 , and putting $a = b\Delta\gamma$ in the differential equation, we obtain two equations for b and a' , by equating to zero the terms independent of, and proportional to $\Delta\gamma$ respectively.

At the separation point b and a' must be equal.

Consider the following points as possible positions of separation:

(i) $x = 1.950$. Put $a = b\Delta\gamma$, where $\Delta\gamma = \gamma - \gamma_0$, γ_0 corresponding to the above value of x . Also

$$\lambda = 0.330, \quad \lambda' = -0.208 - 2.4\Delta\gamma.$$

From the two equations

$$a'^2 = 0.0366, \quad 4.53b^3 - 13.5b^2 + 2.84 = 0,$$

we get

$$a' = -0.191, \quad b = -0.430.$$

(ii) $x = 2.000$, $\gamma = 2.35$, $\lambda = 0.310$, $\lambda' = -0.408 - 2\Delta\gamma$, whence

$$a' = -0.275, \quad b = -0.260.$$

(iii) $x = 2.050$, $\gamma = 2.41$, $\lambda = 0.290$, $\lambda' = -0.525 - 1.65\Delta\gamma$, whence

$$a' = -0.0484, \quad b = -0.271.$$

At the separation point $a' = b$. To find it, we plot the values of a' and b against x (figure 3).

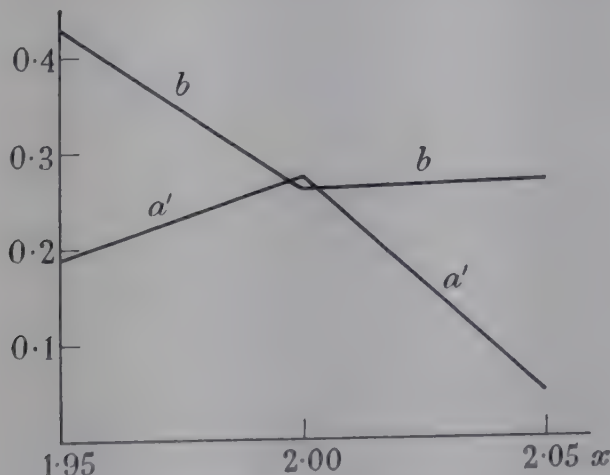


FIGURE 3

As can be seen from figure 3, $a' = b$ at $x = 1.997$ and $x = 2.003$ whence separation is at $x = 2.00$ as against Schubauer's value $x = 1.99 \pm 0.02$, and

$$a' = b = -0.270.$$

7. INTEGRATION DOWNSTREAM OF SEPARATION

We now briefly describe the integration of the equation (3.1) downstream of separation.

The expression (3.7) is an ordinary differential equation in $a(\gamma)$; it is integrated step by step.

We start from the separation point where a vanishes and a' is known. Taking a short interval $\Delta\gamma$ we find the corresponding value of a , namely,

$$a_1 = a'_0 \Delta\gamma,$$

where a'_0 refers to the separation point.

All quantities, except a'_1 and a''_1 , are known at the new point $\Delta\gamma$; putting

$$a''_1 = \frac{a'_1 - a'_0}{\Delta\gamma}$$

we obtain an ordinary equation for a'_1 ; this procedure is then repeated for the next point.

In figures 4 to 6 are plotted the calculated and Schubauer's observed velocities at $x = 2.029$, 2.123 and 2.196 .

The discrepancies between the calculated and observed values arise because we have taken only two terms in the $e^{F(\sigma, \gamma)}$; this most affects the values of u for large y .

The approximate method of integration gives us a general picture of the flow and a satisfactory first approximation for the more rigorous method of integration which will be explained below.

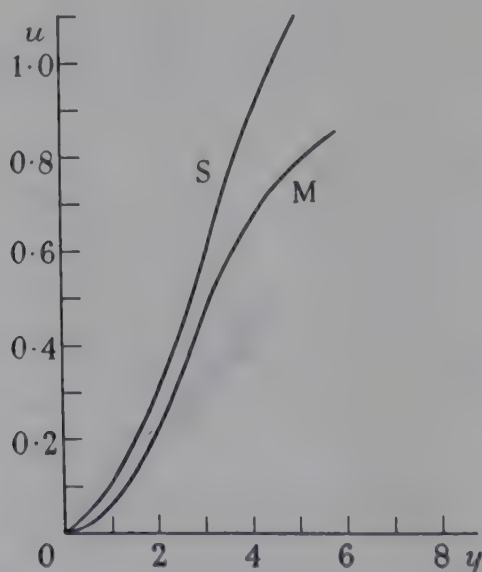


FIGURE 4. $x = 2.029$. S, Schubauer; M, Meksyn.

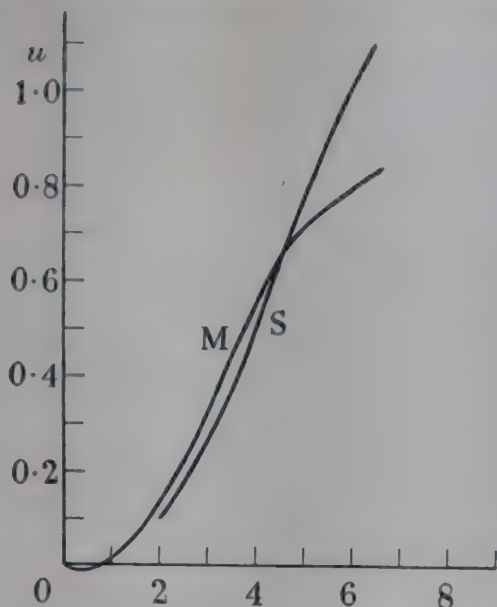


FIGURE 5. $x = 2.123$. S, Schubauer;
M, Meksyn.

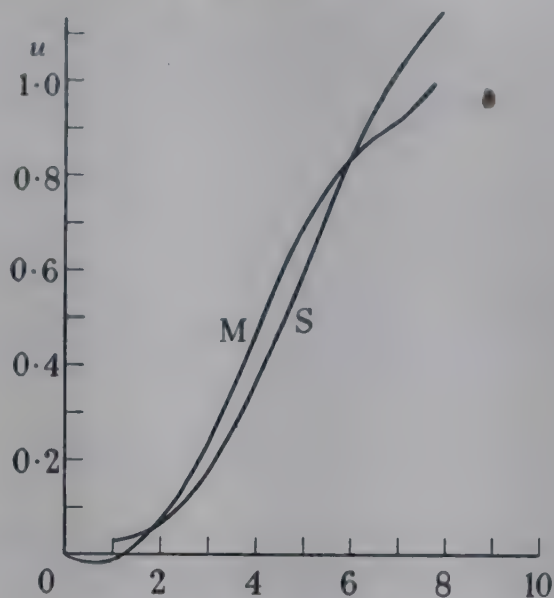


FIGURE 6. $x = 2.196$. S, Schubauer;
M, Meksyn.

8. INTEGRATION OF THE EQUATION UPSTREAM OF SEPARATION. GENERAL METHOD OF INTEGRATION

The approximate solution of the equation

$$f''(\sigma, \gamma) = ae^{-F(\sigma, \gamma)} + e^{-F(\sigma, \gamma)} \left\{ \int_0^\sigma \left[(1 - f'^2) + 2\gamma \left(f' \frac{\partial f'}{\partial \gamma} - \frac{\partial f}{\partial \gamma} f'' \right) \right] e^{F(\sigma, \gamma)} d\sigma \right\}, \quad (8.1)$$

where

$$a = f''(0, \gamma), \quad F(\sigma, \gamma) = \int_0^\sigma f(\sigma, \gamma) d\sigma$$

and

$$f'(\infty, \gamma) = 1 \quad (8.2)$$

in the region of accelerated flow ($\lambda < 0$) presents certain difficulties, because the term $-\lambda\sigma^4/24$ in $-F(\sigma, \gamma)$ is positive, and comparatively large; the convergence of the expression (3.5) is poor, unless many terms are taken, which makes the procedure impracticable.

We shall indicate now a more general method of integration. The equations (8.1) and (8.2) are solved by step-by-step integration.

At the front part of the cylinder the terms in (8.1) proportional to γ are unimportant, and (8.1) is reduced to the well-known equation integrated numerically by Hartree (1937).

For instance, in the case of Schubauer's measurements, the value of $a(\gamma)$ at $x = 0.180$ ($\lambda = -0.456$) obtained from the simplified equation is almost in complete agreement with the value used by Hartree (1939).

Let us assume, to fix ideas, that starting from $x = 0.180$ we have to make use of the complete equation (8.1).

Since the additional terms give only a slight correction we can insert in the right-hand side of (8.1) and in $F(\sigma, \gamma)$ the functions as found from the simplified equation, and, thus, obtain a better approximation for a and $f(\sigma, \gamma)$.

Suppose now that we know the values of $f(\sigma, \gamma)$ and its derivatives with respect to σ at two points $\gamma - \Delta\gamma$ and γ , where $\Delta\gamma$ is a sufficiently small interval.

Since, generally, $f(\sigma, \gamma)$ changes slowly with respect to γ we can put

$$\frac{\partial f}{\partial \gamma} \sim \frac{f(\sigma, \gamma) - f(\sigma, \gamma - \Delta\gamma)}{\Delta\gamma}.$$

To integrate now (8.1), (8.2) for $\gamma + \Delta\gamma$, we proceed as follows.

We assume that $\partial f/\partial \gamma$, $\partial f'/\partial \gamma$, ... do not change within the interval $\gamma - \Delta\gamma$, $\gamma + \Delta\gamma$; we can, thus, find the values of f , f' , etc., at the point $\gamma + \Delta\gamma$; for instance

$$f(\sigma, \gamma + \Delta\gamma) \sim f(\sigma, \gamma) + \Delta\gamma \frac{\partial f(\sigma, \gamma)}{\partial \gamma},$$

and, inserting these expressions in the right-hand side of (8.1) and in $F(\sigma, \gamma)$, we obtain $a(\gamma)$ and $f(\sigma, \gamma)$ at the new point.

By the same procedure we find $f(\sigma, \gamma)$ at the next points.

We shall not make use of this more difficult method of integration, but solve the equations (3.5) and (3.7), for the retarded part of the flow ($\lambda > 0$) by the approximate method sketched in the previous section.

For $x = 1.350$ the value of a was taken from Hartree's (1939) numerical solution, namely, $a = 0.675$, and, plotting the curve of a , we find $a' = -0.175$ at $x = 1.250$.

Making use of (3.5) and (3.7) we then find, as explained above, the value of a' at the point $x = 1.350$ etc.

In figure 7, $a(\gamma)$ is plotted against x , together with Hartree's (1939) results.

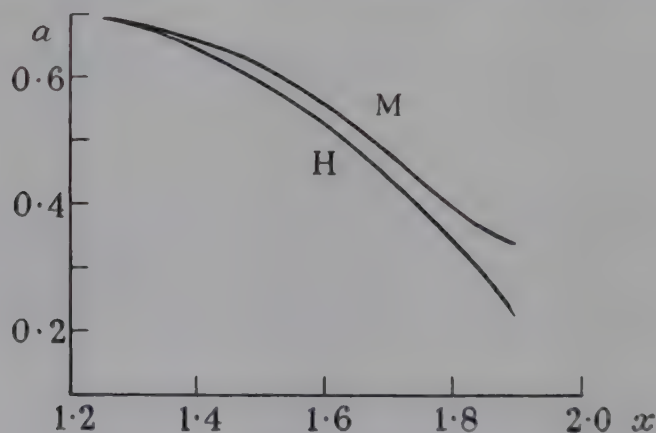


FIGURE 7. H, Hartree; M, Meksyn.

As can be seen, our curve runs somewhat above Hartree's up to $x = 1.832$, and then they diverge.

At this point $a(\gamma)$ begins rapidly to decrease, and the short-cut method of integration is no more applicable.

REFERENCES

- Hartree, D. R. 1937 *Proc. Camb. Phil. Soc.* **33**, 223.
Hartree, D. R. 1939 *Rep. Mem. Aero. Res. Comm.*, London, no. 2427.
Meksyn, D. 1948 *Proc. Roy. Soc. A*, **192**, 545.
Schubauer, G. B. 1935 *N.A.C.A. Tech. Rep.* no. 527.

Integration of the laminar boundary layer equation

II. Retarded flow along a semi-infinite plane

By D. MEKSYN,† *Department of Mathematics, King's College, London*

(Communicated by G. Temple, F.R.S.—Received 25 October 1949)

The equation of the boundary layer for an arbitrary velocity distribution is derived, and applied to a retarded flow along a semi-infinite plane.

Separation can only take place if the function $a(x) = f''(0, x)$ has a singularity at separation of the form $a \sim (x_0 - x)^{\frac{1}{2}}$.

The results of the approximate simple method at integration are compared with Howarth's (1938) calculations; the agreement is rather good.

1. THE BOUNDARY LAYER EQUATION

Consider the case of a semi-infinite plane at rest lying along the stream.

Let the velocity of the incoming flow be in the x direction; y is normal to the plane, and u, v are the corresponding velocities.

The equations of motion in the boundary layer are

$$\left. \begin{aligned} u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \nabla^2 u, \\ \frac{\partial p}{\partial y} &= \text{const.}, \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0. \end{aligned} \right\} \quad (1.1)$$

Denoting by $h(x)$ the velocity outside the boundary layer corresponding to speed unity at infinity, let

$$u = h\bar{u}, \quad v = h\bar{v}, \quad (1.2)$$

where $\bar{u}$ and $\bar{v}$ are functions of y and x .

Let u_1 be the velocity at infinity; the x velocity outside the boundary layer is equal to hu_1 , whence

$$u_1^2 h dh + \frac{1}{\rho} dp = 0. \quad (1.3)$$

From (1.1), (1.2) and (1.3) we obtain

$$\left. \begin{aligned} \bar{u} \frac{\partial \bar{u}}{\partial x} + \bar{v} \frac{\partial \bar{u}}{\partial y} &= -\frac{\partial \log h}{\partial x} (\bar{u}^2 - u_1^2) + \frac{\nu}{h} \frac{\partial^2 \bar{u}}{\partial y^2}, \\ \frac{\partial \bar{u}}{\partial x} + \frac{\partial \bar{v}}{\partial y} + \bar{u} \frac{\partial \log h}{\partial x} &= 0. \end{aligned} \right\} \quad (1.4)$$

Following Emmons & Brainerd (1941) for compressible flow, let

$$\left. \begin{aligned} U &= \frac{\bar{u}}{u_1}, \quad V = \bar{v} \left(\frac{x}{\nu u_1} \right)^{\frac{1}{2}}, \\ \eta &= \left(\frac{u_1}{\nu x} \right)^{\frac{1}{2}} y, \quad \eta U - 2V = \xi. \end{aligned} \right\} \quad (1.5)$$

† Now at the Department of Mathematics, Imperial College of Science and Technology.

Substituting (1.5) in (1.4) we obtain

$$\left. \begin{aligned} \frac{1}{2}u'\xi + \frac{u''}{h} &= -(1-U^2)x \frac{\partial \log h}{\partial x} + xU \frac{\partial U}{\partial x}, \\ \frac{1}{2}(\xi' - U) &= Ux \frac{\partial \log h}{\partial x} + x \frac{\partial U}{\partial x}. \end{aligned} \right\} \quad (1.6)$$

Put

$$\int_0^\eta U(\eta, x) d\eta = f(\eta, x), \quad (1.7)$$

whence substituting (1.7) in the second equation (1.6), integrating with respect to η , and eliminating ξ , we obtain

$$f''' + \left(\frac{h}{2} + x \frac{dh}{dx} \right) ff'' = -(1-f'^2)x \frac{dh}{dx} + hx \left[f' \frac{\partial f'}{\partial x} - \frac{\partial f}{\partial x} f'' \right], \quad (1.8)$$

where primes denote differentiation with respect to η .

The equation (1.8) can be applied to the boundary layer of a curved profile, provided that the co-ordinate x is measured along the boundary from the forward stagnation point, and y is normal to the boundary.

The equation (1.8) is to be integrated under the boundary conditions

$$\left. \begin{aligned} f = f' &= 0 & \text{at } \eta = 0, \\ f' &= 1 & \text{at } \eta = \infty. \end{aligned} \right\} \quad (1.9)$$

Consider now the particular case of retarded flow

$$h = 1 - \beta x. \quad (1.10)$$

A glance at the equation (1.8) shows that its form is preserved if we put

$$\beta x = x^*. \quad (1.11)$$

Taking x^* as a new variable we obtain after few transformations

$$f''' + \frac{1-3x}{2} ff'' = -(1-f'^2)x \frac{dh}{dx} + hx \left[f' \frac{\partial f'}{\partial x} - \frac{\partial f}{\partial x} f'' \right], \quad (1.12)$$

where the star at the x^* was discarded as unnecessary.

The equation (1.12), under the boundary conditions (1.9), can be integrated, as in the previous case (I, § 3), and we obtain

$$\begin{aligned} & \frac{\partial^2 f}{\partial \eta^2}(\eta, x) \\ &= \left\{ a + x \left[\eta - \frac{a^2}{3} \eta^3 + \frac{a(1-6x)}{4} \eta^4 + \frac{x(1-6x)}{5!} \eta^5 + \frac{-\left(\frac{3-x}{2}\right)a^3 + hx a^2 a'}{72} \right] \right. \\ & \quad + hx \left[\frac{aa'}{6} \eta^3 + \frac{a\eta^4}{12} + \frac{x\eta^5}{60} + \frac{25-19x}{2} a^2 a' - (x-x^2) aa'^2 - 2a^3 + 4(x-x^2) a^2 a'' \right. \\ & \quad \left. \left. + \frac{x(10-14x)aa' + (42.5-22x)a^2 + 9x^2(1-x)a'^2 + 9x^2(1-x)aa''}{7!} \eta^7 + \dots \right] \right\} \\ & \quad \times \exp \left\{ r \left[-\frac{a\eta^3}{3!} - \frac{x\eta^4}{4!} - \frac{-\left(\frac{1+x}{2}\right)a^2 + hxaa'}{6!} \eta^6 + \frac{2x^2 a \eta^7}{7!} + \dots \right] \right\}, \quad (1.13) \end{aligned}$$

where

$$r = \frac{1 - 3x}{2},$$

where primes at a denote differentiation with respect to x .

The boundary condition

$$f'(\infty, x) = 1 \quad (1.14)$$

leads to an ordinary differential equation with respect to a which can be integrated step by step.

2. SEPARATION

If in (1.12) the terms depending on the partial derivative with respect to x are disregarded, separation corresponding to the simplified equation will take place at $x = 0.077$.

Since in (1.13) the term aa' is negative, the curve of $a(x)$ will run in the present case above the curve corresponding to the simplified equation, and separation will take place for $x > 0.077$.

We assume, on physical grounds, that at separation all derivatives of a are finite. A glance at (1.13) shows that since $a = 0$, there will be no separation for $\lambda > 0.077$.

It can only take place if aa' does not vanish and is negative right up to and including the separation point, whence

$$a \sim (x_s - x)^{\frac{1}{2}}, \quad (2.1)$$

a result assumed and investigated by Goldstein (1948).

It follows, therefore, that in the present case there can only be a mathematical, but not a physical separation, since on physical grounds the derivatives must be finite.

It does not follow, however, that this case has no physical significance.

In the actual flow the velocity distribution will be slightly modified close to the mathematical separation, and will, thus, lead to a finite solution.

Goldstein (1948) has also shown that there is no real integral after separation which can be joined with the solution at separation.

From this result the conclusion need not be drawn that the boundary-layer equation (1.12) is not valid after separation, but rather that the assumed form of h , namely that the x velocity continuous to decrease after separation is not correct.

A glance at Schubauer's experimental results (I, figure 1) shows that λ decreases after separation.

3. INTEGRATION OF THE EQUATION

In integrating the equations (1.13) and (1.14) only the first two terms in the exponential were retained.

Putting $x = 0$ in (1.12) we reduce it to Blasius's case, whence $a = 0.330$.

A few details of these computations are now given.

Consider the point $x = 0.0125$. In (1.13) we put

$$a' = \frac{a - 0.330}{0.0125}, \quad a'' = 0.$$

We have to assume some value for a in order to evaluate the exponential; it was tentatively assumed that $a = 0.310$.

Making use of a suitable table of integrals we find $a = 0.318$, whence $a' = -1.85$. The procedure for finding a is more troublesome than that for finding a' , since a enters in the exponential.

We accordingly assume that the above value of a' holds good for the next interval. Consider next the point $x = 0.0250$; we obtain $a = 0.318 - 0.023 = 0.295$, and making use of (1.13) we find $a' = -1.90$, whence for $x = 0.0375$, $a = 0.271$.

This procedure can be continued up to $x = 0.110$; for larger values of x there is no real solution, and we should have to make use of the more rigorous procedure which was explained in part I.

At $x = 0.110$, $a = 0.078$ and $a' = -5.8$ whence separation should be at most at
$$x = 0.110 + \frac{0.078}{5.8} = 0.1235,$$

as against Howarth's value $x = 0.120$.

In tables 1 and 2 are given my results and Howarth's. As can be seen the agreement is very satisfactory.

TABLE 1

x	0	0.0125	0.0250	0.0375	0.0500	0.0625	0.0750	0.0875	0.1000	0.1125
a (Meksyn)	0.330	0.318	0.295	0.271	0.245	0.215	0.183	0.150	0.113	0.064
a (Howarth)	0.330	0.309	0.287	0.264	0.238	0.211	0.181	0.149	0.110	0.0615

TABLE 2

η	u/U for $x = 0.0625$		u/U for $x = 0.0875$	
	Meksyn	Howarth	Meksyn	Howarth
0	0	0	0	0
0.4	0.091	0.089	0.068	0.066
0.8	0.191	0.188	0.149	0.146
1.2	0.298	0.293	0.240	0.237
1.6	0.408	0.403	—	—
2.0	0.528	0.513	0.442	0.442
2.4	0.619	0.617	0.544	0.546
2.8	0.711	0.712	0.641	0.646
3.2	0.788	0.794	0.733	0.736
3.6	0.850	0.860	0.819	0.812
4.0	0.904	0.910	0.867	0.872
4.4	0.922	0.946	—	—
4.8	0.949	0.969	0.940	0.951
5.2	0.967	0.984	0.954	0.972
5.6	—	—	0.956	0.985

To integrate the equations the following integrals

$$\int_0^\infty t^n \exp(-\epsilon t^3 - t^4) dt \quad (-1 \leq \epsilon \leq 1),$$
$$\int_0^\infty t^n \exp(-t^3 - \epsilon t^4) dt \quad (0 \leq \epsilon \leq 1),$$

were evaluated for intervals $\Delta\epsilon = 0.1$, and $0 \leq n \leq 7$. For the sake of brevity these results are not given here.

CONCLUSION

No great importance need be attached to the rather good agreement with Howarth's result.

Our analysis, however, shows that a suitable number of terms in the equations (I, (3.5)) and (1.13) can give us a general picture of the flow, and with it an understanding of its intricacies.

It would be important to solve the equation (I, (3.5)) by the rigorous method of step-by-step integration as indicated in (I, § 8). It should not require very laborious computations.

In a considerable part of the flow between the forward stagnation point, and the minimum pressure, the terms proportional to γ can be treated as perturbation terms where, in order to get a better approximation, the function f and its derivatives can be taken from Hartree's (1937) numerical solution.

In the region of retarded flow the expressions are more convergent up to the point where $a(\gamma)$ begins rapidly to change.

The position of separation, in the case of a real flow, can be found as indicated above, and, since the flow changes slowly downstream of separation, the rigorous method of step-by-step integration should give rise to no difficulty, at least for a certain distance downstream.

The integration of the equation of the boundary layer from, and upstream of separation, within the region where $a(\gamma)$ changes rapidly, is a separate and rather difficult problem.

Perhaps the simplest way would be to find an asymptotic solution with a singularity at the separation point by the method developed by Goldstein (1948) in connexion with the problem of a retarded flow along a plane.

I wish to express my thanks to Professor G. Temple, F.R.S., for his interest in my work, and many helpful suggestions; and to the Department of Scientific and Industrial Research for a grant which enabled me to carry out this work.

REFERENCES

- Emmons, H. W. & Brainerd, J. G. 1941 *J. Appl. Mech.* **8**, A 105.
Goldstein, S. 1948 *Q. J. Mech. Appl. Math.* p. 43.
Howarth, L. 1938 *Proc. Roy. Soc. A*, **164**, 547.

The interaction representation in the quantum theory of fields

By J. S. DE WET, *Balliol College, University of Oxford*

(Communicated by R. P. Bell, F.R.S.—Received 9 November 1949)

The interaction representation has recently been introduced into the quantum theory of fields by Tomonaga and Schwinger. Applications of the theory to interacting meson-photon fields have led to apparent difficulties in determining invariant interaction Hamiltonians. Another troublesome feature is the necessity of verifying the integrability conditions of the so-called generalized Schrödinger equation. In the present paper the theory of the interaction representation is presented from a different point of view. It is shown that if two field operators with the same transformation character satisfy two different field equations, there is a unique unitary transformation connecting the field variables on any space-like surface given such a correspondence on one given space-like surface. A differential equation for determining this unique unitary transformation is found which is the analogue of Tomonaga's generalized Schrödinger equation. This gives directly and simply an invariant interaction Hamiltonian and renders unnecessary the explicit verification of the integrability of the Schrödinger equation, since this is known to have a unique solution. To illustrate the simplification introduced by the present theory, the interaction Hamiltonian for the interacting scalar meson-photon fields is calculated. The result is the same as that obtained by Kanesawa & Tomonaga, but it is obtained by a straightforward calculation without the need to add terms to make the Hamiltonian an invariant.

INTRODUCTION

The interaction representation has recently been introduced in the quantum theory of fields by Tomonaga (1946) and by Schwinger (1948). In this representation, instead of having a fixed state vector and field operators satisfying the differential equations of the interacting fields, as in the Heisenberg-Pauli theory, one has a varying state vector, so chosen that the field equations leading to the same expectation values of the operators are the differential equations of the free (non-interacting) fields. In this theory the state vector is a functional of space-like surfaces and is determined in terms of a given state vector on a given surface by a functional differential equation known as the generalized Schrödinger equation. In the way the theory is formulated by Tomonaga, as a generalization of the many-time formalism of Dirac, Fock and Podolsky, an integrability condition of the generalized Schrödinger equation appears, which is the analogue of a similar condition in the many-time theory (Wentzel 1943, p. 134). The role of the integrability condition is to ensure that the Schrödinger equation has a unique solution. When the interaction representation was applied to the interacting photon-electron fields it was found that this integrability condition is satisfied. Kanesawa & Tomonaga (1948*a, b*) and Miyamoto (1948) then considered the applications to the photon field interacting with the scalar and vector-meson fields and found that in these cases the integrability condition was not satisfied by the interaction Hamiltonians they found, nor were these Hamiltonians invariants as one would expect. They showed, however, that it was possible to find additional terms which, when added to their original interaction Hamiltonians, produced invariant Hamiltonians which at the same time did satisfy the integrability condition. Subsequently Belinfante (1949)

applied the interaction representation to the neutral vector-meson field interacting with a field of Dirac particles. He finds an interaction Hamiltonian which is an invariant (without the need for adding any terms) and verifies that this Hamiltonian makes the Schrödinger equation integrable.

Clearly there is something at fault with the work of Kanesawa, Tomonaga and Miyomoto quoted above, for a supposedly relativistically invariant theory, correctly applied, cannot lead to results (in this case interaction Hamiltonians) which are not independent of the choice of the co-ordinate system. This raises the question of how the interaction Hamiltonian is to be calculated correctly. A further point arises from the work of Belinfante. Is it accidental that the interaction Hamiltonian he derives makes the Schrödinger equation integrable, or does it follow from the general theory that the correct Hamiltonian necessarily leads to an integrable Schrödinger equation? If this were to be the case it would render unnecessary the laborious necessity of verifying the integrability.

These are the points considered in the present paper. The theory is formulated in general terms and the approach differs from that of the Japanese workers. Here we show that if two field operators with the same transformation character with respect to transformations of co-ordinates satisfy different field equations derived from two different invariant Lagrangians, and if the field operators are connected by a given unitary transformation on one given space-like surface, a unique unitary transformation between the field operators on any other space-like surface is determined by their respective field equations. It is then shown how to determine this unitary transformation associated with the second surface in terms of the given unitary transformation on the first surface. This is done by choosing a one-parameter family of space-like surfaces containing the two surfaces as members, choosing a co-ordinate system in which the surfaces become $x^4 = \text{constant}$ and then finding a differential equation which determines the unitary transformation as a function of x^4 . This is the analogue of the generalized Schrödinger equation of Tomonaga, and an invariant transformation or interaction Hamiltonian appears quite naturally in the theory. The formalism used in this work is that previously used by the author in proving the general relativistic invariance of quantized field theories (de Wet 1948, 1950). It consists essentially in using general curvilinear co-ordinates in space-time with covariant derivatives appearing in the field equations in the place of ordinary derivatives. One can then, for general families of space-like surfaces, introduce co-ordinate systems in which these surfaces are the surfaces $x^4 = \text{constant}$. In this co-ordinate system the quantized field equations can be written in the usual Heisenberg form with certain additional terms involving Christoffel symbols. All the desired results follow quite simply with the help of this formalism. In the papers quoted above, details of the procedure were not given although the relevant results were quoted. For that reason the complete theory is given in the present paper which is thus complete in itself. We consider here only field equations derived from first-order Lagrangians, i.e. Lagrangians containing first derivatives of the field quantities only. It will be clear that precisely similar results hold for the more general theories derived from higher order Lagrangians. This is the case since the Hamiltonian theory developed here applies also to such equations (see de Wet 1948).

To summarize, the present paper shows that for two fields of the same transformation character, but satisfying different field equations, there is a unique unitary transformation associated with each space-like surface and connecting the two fields on the surface. The theory leads simply to a differential equation for determining this unitary transformation, this being the analogue of Tomonaga's Schrödinger equations, and this equation contains an invariant transformation or interaction Hamiltonian. There is no need to verify the integrability of this equation (i.e. the independence of the solution of the choice of the family of surfaces linking the original surface on which the unitary transformation is given and the final surface on which it is to be determined), since the unitary transformation associated with the final surface is known to be unique and therefore the unitary operator generating the transformation is at most undetermined up to a factor of modulus one. Finally, there is no essential difference as far as the interaction representation is concerned between the interacting photon-electron fields and the interacting photon-meson fields, as was believed by Tomonaga. The interaction Hamiltonian is as easily calculated for the latter case as for the former. To illustrate this the interaction Hamiltonian for the interacting scalar meson-photon fields is calculated here according to the present theory. A simple calculation leads simply and directly to the same result as was found by Kaneshawa & Tomonaga (1948*a*) after resorting to the expedient of adding terms and verifying the integrability condition. In the light of the present theory, therefore, these difficulties are eliminated. The simplification in the calculation of interaction Hamiltonians in the more difficult cases that have been considered is even more striking.

§ 1 deals with the Hamiltonian formulation of the classical field equations. § 2 is concerned with the quantization of these equations and the Heisenberg form of the equations of motion in the appropriate co-ordinate system. In § 3 are considered the interaction representation and how the interaction Hamiltonian is to be determined. To save writing and so as not to confuse the issue, the theory is expounded for the case of tensor fields only, quantized according to Einstein-Bose statistics. At the end of § 3 it is shown briefly that the same results hold when some or all the fields are spinor fields and when some or all the fields are quantized according to Fermi-Dirac statistics. Finally, in § 4 the theory is applied to the determination of the interaction Hamiltonian for the case of the interacting scalar meson-photon fields.

1. THE HAMILTONIAN FORM OF THE CLASSICAL FIELD EQUATIONS

In this section it is shown how the classical field equations

$$\frac{\partial L}{\partial z^\alpha} - \left(\frac{\partial L}{\partial z^\alpha_{,\sigma}} \right)_{,\sigma} = 0 \quad (1.1)$$

can be put into the Hamiltonian form appropriate to a given one-parameter family of space-like surfaces. The field quantities z^α are assumed to be tensor fields with transformation character indicated by a single superscript α . *Covariant differentiation* is denoted by the usual comma notation. Thus $z^\alpha_{,\sigma}$ means the covariant σ derivative of z^α and $(\partial L / \partial z^\alpha_{,\sigma})_{,\sigma}$ the covariant divergence of $\partial L / \partial z^\alpha_{,\sigma}$, this latter symbol meaning

the *ordinary* derivative of L with respect to $z^\alpha_{,\sigma}$. The Lagrangian L is assumed to be an invariant function of z^α , $z^\alpha_{,\sigma}$ and the metric tensor $g_{\mu\nu}$.

Suppose now that $S(\tau)$ is any one-parameter family of space-like surfaces, no two members of which intersect and such that there is one member of the family through every point of space-time. Associated with this family of surfaces we define four vector fields N^σ and $M^\sigma(i)$ ($i = 1, 2, 3$), as follows: $N^\sigma(x^\rho)$ is the unit normal at x^ρ to the surface of the family through x^ρ while $M^\sigma(i)$ ($i = 1, 2, 3$) are three linearly independent vectors orthogonal to N^σ . The vector field N^σ is of course completely determined by the family of surfaces, while the vector fields $M^\sigma(i)$ are to a large extent arbitrary. Let us define a scalar T associated with the family of surfaces by

$$T = T^\sigma_\tau N_\sigma N^\tau, \quad (1.2)$$

where

$$T^\sigma_\tau = L\delta^\sigma_\tau - \frac{\partial L}{\partial z^\alpha_{,\sigma}} z^\alpha_{,\tau}. \quad (1.3)$$

T^σ_τ is the canonical energy-momentum tensor, and we will call T the Hamiltonian function of the Lagrangian associated with the family $S(\tau)$. We define new field variables associated with the family of surfaces by means of the relations

$$\pi_\alpha = \frac{\partial L}{\partial z^\alpha_{,\sigma}} N_\sigma, \quad \theta^\alpha(i) = M^\sigma(i) z^\alpha_{,\sigma}. \quad (1.4)$$

It is clear that $\theta^\alpha(i)$ has the same transformation character as z^α , while π_α transforms contragrediently to z^α . We now show that the field equations (1.1) are equivalent to

$$(N^\sigma \pi_\alpha)_{,\sigma} = \frac{\partial T}{\partial z^\alpha} - \sum_{i=1}^3 \left\{ \frac{\partial T}{\partial \theta^\alpha(i)} M^\sigma(i) \right\}_{,\sigma}, \quad (1.5)$$

where T in (1.5) has been expressed as a function of π_α , $\theta^\alpha(i)$ and z^α , by solving for $z^\alpha_{,\sigma}$ from the four equations (1.4). To prove this equivalence we consider the form of the equations (1.5) in a suitably chosen co-ordinate system and show them to be the same as (1.1) in this co-ordinate system. The special co-ordinate system, which we will call a *natural co-ordinate system* associated with the family of surfaces $S(\tau)$, is one with the following properties: (i) the family of surfaces $S(\tau)$ become the family $x^4 = \text{constant}$; (ii) the components g^{4i} and g_{4i} ($i = 1, 2, 3$) of the metric tensor all vanish in this co-ordinate system. It is easily seen that such co-ordinate systems do exist. Furthermore, there is a certain amount of arbitrariness left. Given any natural co-ordinate system x^σ , the transformation $\bar{x}^4 = f(x^4)$; $\bar{x}^i = g^i(x^1, x^2, x^3)$, produces another natural co-ordinate system.

In a natural co-ordinate system the various tensor fields we have introduced above have the following components:

$$\left. \begin{aligned} N^\sigma &= \sqrt{(g^{44})} \delta^\sigma_4, & N_\sigma &= \frac{1}{\sqrt{(g^{44})}} \delta^4_\sigma, & M^4(i) &= 0 \quad (i = 1, 2, 3), \\ \pi_\alpha &= \frac{1}{\sqrt{(g^{44})}} \frac{\partial L}{\partial z^\alpha_{,4}}, & \theta^\alpha(i) &= M^j(i) z^\alpha_{,j} \quad (j \text{ summed over } 1, 2, 3 \text{ only}), \\ T &= L - \sqrt{(g^{44})} \pi_\alpha z^\alpha_{,4}. \end{aligned} \right\} \quad (1.6)$$

We now write out the equations (1.5) in full and then substitute the appropriate components of the tensor quantities in the natural co-ordinate system as given by (1.6). Before we do this we introduce a new convention to save writing. If $A_{\alpha_1 \dots \alpha_r}^\sigma$ is any tensor field of transformation character indicated by the superscript and subscripts

$$(A_{\alpha_1 \dots \alpha_r}^\sigma)_{,\sigma} = \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^\sigma} \{ \sqrt{(-g)} A_{\alpha_1 \dots \alpha_r}^\sigma \} - A_{\beta \alpha_2 \dots \alpha_r}^\sigma \Gamma_{\sigma \alpha_1}^\beta - \dots - A_{\alpha_1 \dots \alpha_{r-1} \beta}^\sigma \Gamma_{\sigma r}^\beta, \quad (1.7)$$

where $\Gamma_{\alpha\sigma}^\beta$ is the usual Christoffel symbol of the second kind (the notation used is that of Veblen (1933, p. 34)). In this work a single subscript is used to indicate no, or one or more tensor indices. We will therefore write

$$(A_\alpha^\sigma)_{,\sigma} = \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^\alpha} \{ \sqrt{(-g)} A_\alpha^\sigma \} - [A_\beta^\sigma \Gamma_{\sigma\alpha}^\beta], \quad (1.8)$$

where the square brackets on the right-hand side of (1.8) are to mean a sum of terms involving Christoffel symbols like on the right-hand side of (1.7), the number of terms in the sum being equal to the number of tensor indices represented by the single subscript α . For a contravariant index α representing any number of contravariant indices, we write likewise

$$(A^{\sigma\alpha})_{,\sigma} = \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^\sigma} \{ \sqrt{(-g)} A^{\sigma\alpha} \} + [A^{\sigma\beta} \Gamma_{\beta\sigma}^\alpha]. \quad (1.9)$$

With this notation (1.5) can be written as

$$\begin{aligned} & \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^\sigma} \{ \sqrt{(-g)} N^\sigma \pi_\alpha \} - [N^\sigma \pi_\beta \Gamma_{\sigma\alpha}^\beta] \\ &= \frac{\partial T}{\partial z^\alpha} - \sum_{i=1}^3 \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^\sigma} \left\{ \frac{\partial T}{\partial \theta^\alpha(i)} M^\sigma(i) \sqrt{(-g)} \right\} + \sum_{i=1}^3 \left[\frac{\partial T}{\partial \theta^\beta(i)} M^\sigma(i) \Gamma_{\sigma\alpha}^\beta \right]. \end{aligned} \quad (1.10)$$

In a natural co-ordinate system, this becomes (using (1.6))

$$\begin{aligned} & \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^4} \{ \sqrt{(-g)} \sqrt{(g^{44})} \pi_\alpha \} - [\sqrt{(g^{44})} \pi_\beta \Gamma_{\sigma\alpha}^\beta] \\ &= \frac{\partial T}{\partial z^\alpha} - \sum_{i=1}^3 \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^j} \left\{ \frac{\partial T}{\partial \theta^\alpha(i)} M^j(i) \sqrt{(-g)} \right\} + \sum_{i=1}^3 \left[\frac{\partial T}{\partial \theta^\beta(i)} M^j(i) \Gamma_{j\alpha}^\beta \right], \end{aligned} \quad (1.11)$$

where in the right-hand side of (1.11) j is summed over 1, 2, 3 only. Now, in the natural co-ordinate system $T = L - \sqrt{(g^{44})} \pi_\alpha z_{,4}^\alpha$, and therefore

$$\frac{\partial T}{\partial z^\alpha} = \frac{\partial L}{\partial z^\alpha} \quad \text{and} \quad \sum_{i=1}^3 \frac{\partial T}{\partial \theta^\alpha(i)} M^j(i) = \frac{\partial T}{\partial z_{,j}^\alpha} = \frac{\partial L}{\partial z_{,j}^\alpha},$$

while from the definition of π_α we have $\sqrt{(g^{44})} \pi_\alpha = \frac{\partial L}{\partial z_{,4}^\alpha}$. Substituting all these in (1.11) we get

$$\begin{aligned} & \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^4} \left\{ \sqrt{(-g)} \frac{\partial L}{\partial z_{,4}^\alpha} \right\} - \left[\frac{\partial L}{\partial z_{,4}^\beta} \Gamma_{\sigma\alpha}^\beta \right] \\ &= \frac{\partial L}{\partial z^\alpha} - \frac{1}{\sqrt{(-g)}} \frac{\partial}{\partial x^j} \left\{ \sqrt{(-g)} \frac{\partial L}{\partial z_{,j}^\alpha} \right\} + \left[\frac{\partial L}{\partial z_{,j}^\beta} \Gamma_{j\alpha}^\beta \right], \end{aligned} \quad (1.12)$$

which are just the field equations (1.1) written out in full and separated into two parts. We have thus established the equivalence of the equations (1.1) and (1.5). We note further that from the definition of T by (1.2) and of π_α by (1.4) it follows that $T = L - \pi_\alpha z_{,\tau}^\alpha N^\tau$, and therefore that

$$z_{,\tau}^\alpha N^\tau = -\frac{\partial T}{\partial \pi_\alpha}. \quad (1.13)$$

(1.13) is thus an alternative relation to the first of (1.4). The result we have thus proved is that the equations (1.1) are equivalent to

$$\left. \begin{aligned} (N^\sigma \pi_\alpha)_{,\sigma} &= \frac{\partial T}{\partial z^\alpha} - \sum_{i=1}^3 \left\{ \frac{\partial T}{\partial \theta^\alpha(i)} M^\sigma(i) \right\}_{,\sigma}, \\ z_{,\sigma}^\alpha N^\sigma &= -\frac{\partial T}{\partial \pi_\alpha}, \quad z_{,\sigma}^\alpha M^\sigma(i) = \theta^\alpha(i). \end{aligned} \right\} \quad (1.14)$$

Before completing this section we write the equations (1.14) in a slightly more convenient form, in a natural co-ordinate system. Defining P_α and $\mathcal{H}$ by

$$P_\alpha = \sqrt{(-g)} \sqrt{(g^{44})} \pi_\alpha, \quad \mathcal{H} = \sqrt{(-g)} T, \quad (1.15)$$

it follows from (1.11) that the equations (1.14) in a natural co-ordinate system can be written as

$$\left. \begin{aligned} \frac{\partial P_\alpha}{\partial x^4} - [P_\beta \Gamma_{\alpha 4}^\beta] &= \frac{\partial \mathcal{H}}{\partial z^\alpha} - \sum_{i=1}^3 \frac{\partial}{\partial x_j} \left\{ \frac{\partial \mathcal{H}}{\partial \theta^\alpha(i)} M^j(i) \right\} + \sum_{i=1}^3 \left[\frac{\partial \mathcal{H}}{\partial \theta^\beta(i)} M^j(i) \Gamma_{\alpha j}^\beta \right], \\ \frac{\partial z^\alpha}{\partial x^4} + [z^\beta \Gamma_{\beta 4}^\alpha] &= -\frac{\partial \mathcal{H}}{\partial P_\alpha}, \\ z_{,j}^\alpha M^j(i) &= \theta^\alpha(i). \end{aligned} \right\} \quad (1.16)$$

2. THE QUANTIZED FIELD

In this section we concern ourselves with the quantized field equations and in particular with showing how, for a given one-parameter family of space-like surfaces and an associated natural co-ordinate system, the quantized field equations can be written in a form analogous to the usual Heisenberg equations of motion. We consider only the Einstein-Bose quantization, as the Fermi-Dirac quantization leads to similar results.

As is well known, the quantization is effected by regarding the field variables as operators satisfying certain commutation relations. The order of the terms in the various expressions become of importance and care has to be exercised. In particular, in forming the operator T defined by (1.2) the order of terms is chosen so as to make T self-adjoint. In the Einstein-Bose quantized field the commutation relations between the operators associated with the same member of any family of space-like surfaces are (in a natural co-ordinate system)

$$\left. \begin{aligned} [P_\alpha(\mathbf{x}; x^4), P_\beta(\mathbf{x}', x^4)] &= 0, \\ [z^\alpha(\mathbf{x}, x^4), z^\beta(\mathbf{x}', x^4)] &= 0, \\ [P_\alpha(\mathbf{x}, x^4), z^\beta(\mathbf{x}', x^4)] &= i\hbar \delta_\alpha^\beta \delta(\mathbf{x} - \mathbf{x}'), \end{aligned} \right\} \quad (2.1)$$

where $[A, B] = AB - BA$, $\mathbf{x}$ stands for $x^1 x^2 x^3$ and

$$\delta(\mathbf{x} - \mathbf{x}') = \delta(x^1 - x'^1) \delta(x^2 - x'^2) \delta(x^3 - x'^3),$$

δ being the usual Dirac δ function. Since the solution of the field equations (1.1) is completely determined if P_α and z^α are given on any one space-like surface, it follows that given commutation relations like (2.1) for the surface $x^4 = 0$, the commutation relations for the other surfaces $x^4 = \text{constant}$ are determined by the field equations. Consequently, it is necessary to show that the assumed relations are consistent with the field equations. As is easily seen, this consistence is equivalent to the so-called general relativistic invariance of the quantization, and this has recently been proved very simply by the author using the present technique (de Wet 1950).

We now show that in virtue of the commutation relations (2.1) the field equations can be written in the natural co-ordinate system as

$$\left. \begin{aligned} i\hbar \left\{ \frac{\partial P_\alpha}{\partial x^4} - [P_\beta \Gamma_{\alpha 4}^\beta] \right\} &= [P_\alpha(\mathbf{x}; x^4), H(x^4)], \\ i\hbar \left\{ \frac{\partial z^\alpha}{\partial x^4} + [z^\beta \Gamma_{\beta 4}^\alpha] \right\} &= [z^\alpha(\mathbf{x}; x^4), H(x^4)]. \end{aligned} \right\} \quad (2.2)$$

where H is a function of x^4 only defined by

$$H = \iiint \mathcal{H} dx^1 dx^2 dx^3, \quad (2.3)$$

$\mathcal{H}$ being defined by (1.17) and the integral being over the space-like surface of the family labelled by the given x^4 . To prove (2.2) we note that

$$\begin{aligned} [P_\alpha(\mathbf{x}), H] &= \iiint [P_\alpha(\mathbf{x}), \mathcal{H}(\mathbf{x}')] dx'^1 dx'^2 dx'^3 \\ &= \iiint \left\{ \frac{\partial \mathcal{H}(\mathbf{x}')}{\partial z^\beta(\mathbf{x}')} [P_\alpha(\mathbf{x}), z^\beta(\mathbf{x}')] + \sum_{i=1}^3 \frac{\partial \mathcal{H}(\mathbf{x}')}{\partial \theta^\beta(i)} [P_\alpha(\mathbf{x}), \theta^\beta(i; \mathbf{x}')] \right\} dx'^1 dx'^2 dx'^3. \end{aligned} \quad (2.4)$$

Now the commutator $[P_\alpha(\mathbf{x}), \theta^\beta(i; \mathbf{x}')] is determined by the last of (2.1), for since $z_{,j}^\beta = \partial z^\beta / \partial x^j + [z^\rho \Gamma_{\rho j}^\beta]$, we have$

$$[P_\alpha(\mathbf{x}), z_{,j}^\beta(\mathbf{x}')] = i\hbar \delta_\alpha^\beta \frac{\partial}{\partial x'^j} \{\delta(\mathbf{x} - \mathbf{x}')\} + i\hbar [\delta_\alpha^\rho \Gamma_{\rho j}^\beta] \delta(\mathbf{x} - \mathbf{x}'), \quad (2.5)$$

and, therefore,

$$[P_\alpha(\mathbf{x}), \theta^\beta(i; \mathbf{x}')] = i\hbar \delta_\alpha^\beta M^j(i) \frac{\partial}{\partial x'^j} \{\delta(\mathbf{x} - \mathbf{x}')\} + i\hbar [\delta_\alpha^\rho \Gamma_{\rho j}^\beta] M^j(i) \delta(\mathbf{x} - \mathbf{x}'). \quad (2.6)$$

Substituting from (2.6) into (2.4) it follows that

$$\begin{aligned} [P_\alpha(\mathbf{x}), H] &= i\hbar \left\{ \frac{\partial \mathcal{H}(x)}{\partial z^\alpha(x)} - \sum_{i=1}^3 \frac{\partial}{\partial x^j} \left(\frac{\partial \mathcal{H}(x)}{\partial \theta^\alpha(i)} M^j(i) \right) + \sum_{i=1}^3 \left[\frac{\partial \mathcal{H}}{\partial \theta^\beta(i)} M^j(i) \Gamma_{\alpha j}^\beta \right] \right\} \\ &= i\hbar \left\{ \frac{\partial P_\alpha}{\partial x^4} - [P_\beta \Gamma_{\alpha 4}^\beta] \right\}, \end{aligned} \quad (2.7)$$

using (1.18). It follows similarly that

$$[z^\alpha(\mathbf{x}), H] = -i\hbar \frac{\partial \mathcal{H}}{\partial \pi_\alpha} = i\hbar \left\{ \frac{\partial z^\alpha}{\partial x^4} + [z^\beta \Gamma_{\beta 4}^\alpha] \right\}. \quad (2.8)$$

We have thus shown that in a natural co-ordinate system associated with a given one-parameter family of space-like surfaces the quantized field equations can be written in the form (2.2).

3. THE INTERACTION REPRESENTATION

Suppose now that we have a set of field operators z^α as before, satisfying the field equations

$$0 = \frac{\partial L_1}{\partial z^\alpha} - \left(\frac{\partial L_1}{\partial z_{,\sigma}^\alpha} \right)_{,\sigma}, \quad (3.1)$$

and another set of field operators $\tilde{z}^\alpha$ with the same transformation character as z^α but satisfying the field equations

$$0 = \frac{\partial L_2}{\partial \tilde{z}^\alpha} - \left(\frac{\partial L_2}{\partial \tilde{z}_{,\sigma}^\alpha} \right)_{,\sigma}, \quad (3.2)$$

where L_1 and L_2 are respectively invariant Lagrangian functions of z^α ; $z_{,\sigma}^\alpha$ and $\tilde{z}^\alpha$; $\tilde{z}_{,\sigma}^\alpha$. If S is any space-like surface we define as before field variables π_α and $\tilde{\pi}_\alpha$ by

$$\pi_\alpha = \frac{\partial L_1}{\partial z_{,\sigma}^\alpha} N_\sigma, \quad \tilde{\pi}_\alpha = \frac{\partial L_2}{\partial \tilde{z}_{,\sigma}^\alpha} N_\sigma,$$

where N_σ is the unit normal as before. Suppose now that S_0 and S_1 are any two non-intersecting space-like surfaces. The variables z^α , etc., on S_0 we denote by $z^\alpha(0)$, $\pi_\alpha(0)$, $\tilde{z}^\alpha(0)$, $\tilde{\pi}_\alpha(0)$ with a similar notation for the variables on S_1 . Now if $z^\alpha(1)$, $\pi_\alpha(1)$ are any given operator functions on S_1 , the field equations (3.1) determine $z^\alpha(0)$ and $\pi_\alpha(0)$ on S_0 uniquely. Let U_0 be a given unitary operator and define $\tilde{z}^\alpha(0)$ and $\tilde{\pi}_\alpha(0)$ on S_0 by the transformation

$$\left. \begin{aligned} \tilde{z}^\alpha(0) &= U_0 z^\alpha(0) U_0^*, \\ \tilde{\pi}_\alpha(0) &= U_0 \pi_\alpha(0) U_0^*. \end{aligned} \right\} \quad (3.3)$$

These values of $\tilde{z}^\alpha(0)$ and $\tilde{\pi}_\alpha(0)$ on S_0 again, in virtue of the field equations (3.2), determine uniquely the operator functions $\tilde{z}^\alpha(1)$ and $\tilde{\pi}_\alpha(1)$ on S_1 . Therefore the field equations (3.1) and (3.2) and a given unitary operator U_0 associated with S_0 determine $\tilde{z}^\alpha(1)$ and $\tilde{\pi}_\alpha(1)$ uniquely in terms of arbitrarily given variables $z^\alpha(1)$ and $\pi_\alpha(1)$ on S_1 . Thus for a given unitary transformation connecting the variables $\tilde{z}^\alpha(0)$, $\tilde{\pi}_\alpha(0)$ and $z^\alpha(0)$, $\pi_\alpha(0)$ on S_0 there is a uniquely determined correspondence associated with S_1 between the variables $\pi_\alpha(1)$, $z^\alpha(1)$ and $\tilde{\pi}_\alpha(1)$, $\tilde{z}^\alpha(1)$. Moreover, this correspondence is a unitary correspondence. For suppose the given operators $\pi_\alpha(1)$ and $z^\alpha(1)$ on S_1 that we started with are such that the corresponding $P_\alpha(1)$ and $z^\alpha(1)$ satisfy commutation relations like (2.1). Then, in virtue of the relativistic invariance of the quantization, similar commutation relations are satisfied by $P_\alpha(0)$ and $z^\alpha(0)$ on S_0 and also by $\tilde{P}_\alpha(0)$ and $\tilde{z}^\alpha(0)$ on S_0 because of (3.3). This again implies similar commutation relations between $\tilde{P}_\alpha(1)$ and $\tilde{z}^\alpha(1)$ on S_1 (again by the relativistic invariance of the quantization). It follows that the correspondence between the variables $\tilde{\pi}_\alpha(1)$, $\tilde{z}^\alpha(1)$ and $\pi_\alpha(1)$, $z^\alpha(1)$ on S_1 is a unitary correspondence (Dirac 1947, p. 106). We have thus

shown that, given a unitary transformation (3.3) between the variables $\pi_\alpha(0)z^\alpha(0)$ and $\tilde{\pi}_\alpha(0), \tilde{z}^\alpha(0)$ on S_0 , the field equations determine a unique unitary transformation

$$\left. \begin{aligned} \tilde{z}^\alpha(1) &= U_1 z^\alpha(1) U_1^*, \\ \tilde{\pi}_\alpha(1) &= U_1 \pi_\alpha(1) U_1^*, \end{aligned} \right\} \quad (3.4)$$

associated with S_1 between the corresponding variables on S_1 . The same result holds even if the two surfaces intersect. For, if S_0 and S_1 intersect, one can choose a third surface S_2 which does not intersect S_0 or S_1 and then apply the previous argument first to S_0 and S_2 and then to S_2 and S_1 . Thus, given a unitary transformation (3.3) between the field operators satisfying field equations (3.1) and (3.2) respectively on one given space-like surface, there is a unique unitary transformation of the same kind associated with every other space-like surface.

It remains to show how to determine the unitary transformation associated with the surface S_1 in terms of the given transformation on S_0 . This is a simple matter as we shall see. Let us consider the case where the field equations (3.1) and (3.2) are quantized according to Einstein-Bose statistics, and the field operators therefore satisfy commutation relations like (2.1) on every space-like surface. Let S_0 and S_1 be two non-intersecting surfaces (as noted above, if we can determine the relation between the unitary transformations of two non-intersecting surfaces, we can at once determine the relation for intersecting surfaces as well). Choose *any* one-parameter family of space-like surfaces including S_0 and S_1 as members. Let $T_1(\pi_\alpha, z^\alpha, \theta^\alpha)$ and $T_2(\tilde{\pi}_\alpha, \tilde{z}^\alpha, \tilde{\theta}^\alpha)$ be the Hamiltonian functions derived from the Lagrangians L_1 and L_2 respectively associated with the family of surfaces. In a natural co-ordinate system the unitary transformation associated with each surface of the family will be generated by a unitary operator $U(x^4)$ which is a function of x^4 only. We now show how to determine a differential equation for $U(x^4)$. For every value of x^4 we have

$$\tilde{z}^\alpha(\mathbf{x}; x^4) = U(x^4) z^\alpha(\mathbf{x}; x^4) U^*(x^4), \quad (3.5)$$

$$\text{and, therefore,} \quad i\hbar \frac{\partial \tilde{z}^\alpha}{\partial x^4} = i\hbar \frac{\partial U}{\partial x^4} z^\alpha U^* + i\hbar U z^\alpha \frac{\partial U^*}{\partial x^4} + U \left(i\hbar \frac{\partial z^\alpha}{\partial x^4} \right) U^*. \quad (3.6)$$

Now from the work of § 2, if the field equations are quantized according to Einstein-Bose statistics,

$$\left. \begin{aligned} i\hbar \frac{\partial z^\alpha}{\partial x^4} &= -i\hbar [z^\beta \Gamma_{\beta 4}^\alpha] + [z^\alpha, H_1], \\ i\hbar \frac{\partial \tilde{z}^\alpha}{\partial x^4} &= -i\hbar [\tilde{z}^\beta \Gamma_{\beta 4}^\alpha] + [\tilde{z}^\alpha, H_2], \end{aligned} \right\} \quad (3.7)$$

$$\text{where} \quad H_1 = \iiint \sqrt{(-g)} T_1 dx^1 dx^2 dx^3, \quad H_2 = \iiint \sqrt{(-g)} T_2 dx^1 dx^2 dx^3.$$

Substituting from (3.7) in (3.6) we get

$$\begin{aligned} i\hbar \frac{\partial U}{\partial x^4} z^\alpha U^* + i\hbar U z^\alpha \frac{\partial U^*}{\partial x^4} &= [\tilde{z}^\alpha, H_2] - U [z^\alpha, H_1] U^* \\ &= [\tilde{z}^\alpha, H_2 - \tilde{H}_1], \end{aligned} \quad (3.8)$$

where

$$\tilde{H}_1 = U H_1 U^*. \quad (3.9)$$

Since $U(x^4)$ is a unitary operator $UU^* = U^*U = 1$, and therefore

$$\frac{\partial U}{\partial x^4} U^* + U \frac{\partial U^*}{\partial x^4} = 0. \quad (3.10)$$

Using this in (3.8) and putting $z^\alpha = U^* \tilde{z}^\alpha U$ we get

$$\left[i\hbar \frac{\partial U}{\partial x^4} U^*, \tilde{z}^\alpha \right] = [\tilde{H}_1 - H_2, \tilde{z}^\alpha]. \quad (3.11)$$

Clearly (3.11) will be satisfied if

$$i\hbar \frac{\partial U}{\partial x^4} U^* = \tilde{H}_1 - H_2. \quad (3.12)$$

Now

$$\begin{aligned} H_1 &= U \left\{ \iiint \sqrt{(-g)} T_1(\pi_\alpha, z^\alpha, \theta^\alpha) dx^1 dx^2 dx^3 \right\} U^* \\ &= \iiint \sqrt{(-g)} T_1(\tilde{\pi}_\alpha, \tilde{z}^\alpha, \tilde{\theta}^\alpha) dx^1 dx^2 dx^3, \end{aligned} \quad (3.13)$$

if T_1 is a polynomial in the field variables as it is in applications. Therefore (3.12) can be written as

$$i\hbar \frac{\partial U}{\partial x^4} U^* = \iiint \sqrt{(-g)} K(\tilde{\pi}_\alpha, \tilde{z}^\alpha, \tilde{\theta}^\alpha) dx^1 dx^2 dx^3, \quad (3.14)$$

where

$$K(\tilde{\pi}_\alpha, \tilde{z}^\alpha, \tilde{\theta}^\alpha) = T_1(\tilde{\pi}_\alpha, \tilde{z}^\alpha, \tilde{\theta}^\alpha) - T_2(\tilde{\pi}_\alpha, \tilde{z}^\alpha, \tilde{\theta}^\alpha). \quad (3.15)$$

The function K , which we shall call the *transformation Hamiltonian* associated with the two Lagrangians and the family of surfaces, is an invariant, being the difference of two invariants. Given that $U = U_0$ on the surface S_0 the equation (3.14) determines a U_1 on S_1 , and this U_1 generates the unique unitary transformation on S_1 . By multiplying (3.14) in front by U^* and behind by U we get an alternative equation to (3.14)

$$i\hbar U^* \frac{\partial U}{\partial x^4} = \iiint \sqrt{(-g)} K(\pi_\alpha, z^\alpha, \theta^\alpha) dx^1 dx^2 dx^3. \quad (3.16)$$

The application of the above work to the so-called interaction representation is immediate. Suppose the field equations satisfied by the field variables z^α are (3.1) with $L_1 = L_0 + \epsilon L_{\text{int.}}$, where L_0 and $L_{\text{int.}}$ are both invariants and ϵ is simply a parameter. z^α we will call the interacting field. Then one seeks a unique unitary transformation associated with every space-like surface in space-time with the property that the transformed field variables $\tilde{z}^\alpha$ defined by $\tilde{z}^\alpha = U z^\alpha U^*$ satisfies the field equations (3.2) with $L_2 = L_0$, i.e. the interaction-free field equations. In order to keep the expectation values of the transformed operators unchanged one associates with each space-like surface a state vector $U\phi_0$, where ϕ_0 is the fixed state vector of the interacting field. Now we have seen from the above work that there is such a unique unitary transformation associated with each space-like surface and that the transformation associated with S_1 is determined from that associated with S_0 by taking any one-parameter family of surfaces linking S_0 and S_1 , choosing the appropriate natural co-ordinate system associated with this family and then deriving the transformation by means of the differential equation (3.15) or (3.16), where K is the transformation Hamiltonian associated with the two Lagrangians, which we

now call the interaction Hamiltonian. To determine this, let $T(\epsilon)$ be the Hamiltonian function (expressed in terms of the variables $\pi_\alpha, z^\alpha, \theta^\alpha$) associated with $L_0 + \epsilon L_{\text{int}}$. Then $T_1 = T(\epsilon)$ and $T_2 = T(0)$, where $T(0)$ is obtained from $T(\epsilon)$ by putting $\epsilon = 0$. The interaction Hamiltonian T_{int} is then given by $T_{\text{int}} = T(\epsilon) - T(0)$.

To complete our discussion, let us consider the uniqueness of the state vector associated with every space-like surface in the interaction representation and show why, in our treatment, the integrability of the so-called generalized Schrödinger equation, which plays an important part in the Tomonaga theory, needs no discussion. We note first that, although the unitary transformation (3.4) associated with each surface is uniquely determined by the surface, this leaves the unitary operator U which generates the transformation undetermined up to a factor of modulus one. If we therefore define the state vector associated with each surface by $\phi = U\phi_0$, ϕ_0 being the fixed state vector in the original representation ϕ will be undetermined to the same extent as U , but this, of course, does not in any way affect expectation values calculated from the state vector.

The analogue in our theory of the generalized Schrödinger equation in the Tomonaga theory is the equation (3.14). The question of the integrability of the Schrödinger equation amounts to the question whether, given $U = U_0$ on S_0 , the equation (3.14) leads to the same U_1 on S_1 irrespective of the way in which the intermediate family of surfaces has been chosen. Since the result of integrating the equation (3.14) through two different intervening families can differ at most by a factor of modulus one which makes no difference at all, it is clear that it becomes unnecessary to discuss the integrability.

We have thus shown how to find the interaction Hamiltonian which by means of (3.14) associates with every space-like surface a unique unitary transformation. When this transformation is applied to the interacting field variables on the surface, the transformed variables satisfy the field equations of the free field. With every space-like surface is also associated a state vector which is completely determined by the surface apart from factors of modulus one. The explicit verification of the integrability of equation (3.14) is not necessary.

It is clear that all the above considerations also apply to any number of tensor fields. Spinor fields also fit quite simply into this framework. For the purposes of general transformations of co-ordinates one simply regards spinors as scalars and the bridge-spinors connecting the spinor fields with tensor quantities one simply treats as quantities with transformation character determined by their tensor indices only. Likewise fields quantized according to Fermi-Dirac statistics lead to exactly the same theory, as is very easily seen.

4. APPLICATION TO THE DETERMINATION OF THE INTERACTION HAMILTONIAN FOR THE INTERACTING SCALAR-MESON PHOTON FIELDS

We now apply the above theory to the problem of finding the interaction Hamiltonian for the interacting scalar-meson photon fields. This has already been found by Kanesawa & Tomonaga (1948a), but only after considerable difficulty. The

purpose of the present section is to show that the theory developed above leads simply and directly to the same result. The Lagrangian of the interacting field is

$$L = \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}(A_{,\mu}^\mu)^2 + g^{\mu\nu}(\phi_{,\mu}^* + \epsilon A_{\mu}\phi^*)(\phi_{,\nu} - \epsilon\phi A_{\nu}) + K^2\phi^*\phi, \quad (4.1)$$

where the photon field variables are the A_μ , the scalar meson field variables are ϕ and ϕ^* , and where $F_{\mu\nu} = A_{\mu,\nu} - A_{\nu,\mu}$ and $\epsilon = ie/\hbar c$. In order to calculate the interaction Hamiltonian associated with a given family of space-like surfaces, we find the Hamiltonian function associated with the Lagrangian (4.1) and subtract the Hamiltonian function associated with the Lagrangian obtained by putting $\epsilon = 0$ in (4.1). The calculation is simplified if we work in the natural co-ordinate system associated with the surfaces. Let us denote by B^μ the variables conjugate to A_μ and by χ and χ^* the variables conjugate to ϕ and ϕ^* . Then

$$B^\mu = \frac{\partial L}{\partial A_{\mu,\nu}} N_\nu, \quad \chi = \frac{\partial L}{\partial \phi_{,\nu}} N_\nu, \quad \chi^* = \frac{\partial L}{\partial \phi_{,\nu}^*} N_\nu. \quad (4.2)$$

In a natural co-ordinate system this becomes

$$B^\mu = [F^{\mu\nu} + g^{\mu\nu}A_{,\rho}^\rho] N_\nu = (F^{\mu 4} + g^{\mu 4}A_{,\rho}^\rho) \frac{1}{\sqrt{(g^{44})}}, \quad (4.3)$$

$$\left. \begin{aligned} \chi &= g^{\mu\nu}(\phi_{,\mu}^* + \epsilon A_{\mu}\phi) N_\nu = \sqrt{(g^{44})}(\phi_{,4}^* + \epsilon A_4\phi^*), \\ \chi^* &= \sqrt{(g^{44})}(\phi_{,4} - \epsilon\phi A_4). \end{aligned} \right\} \quad (4.4)$$

From (4.3) it follows that

$$B_\mu = (F_{\mu 4}g^{44} + \delta_\mu^4 A_{,i}^i + \delta_\mu^4 A_{4,4}g^{44}) \frac{1}{\sqrt{(g^{44})}} \quad (i = 1, 2, 3),$$

and therefore

$$B_i = (A_{i,4} - A_{4,i}) \sqrt{(g^{44})}, \quad B_4 = (A_{,i}^i + A_{4,4}g^{44}) \frac{1}{\sqrt{(g^{44})}}$$

or

$$A_{i,4} = \frac{B_i}{\sqrt{(g^{44})}} + A_{4,i}, \quad A_{4,4} = (B_4 \sqrt{(g^{44})} - A_{,i}^i) \frac{1}{g^{44}}. \quad (4.5)$$

From (4.4) we get

$$\phi_{,4} = \frac{\chi^*}{\sqrt{(g^{44})}} + \epsilon\phi A_4, \quad \phi_{,4}^* = \frac{\chi}{\sqrt{(g^{44})}} - \epsilon A_4\phi^*. \quad (4.6)$$

Defining the Hamiltonian function $T(\epsilon)$ corresponding to the Lagrangian L of (4.1) by

$$T(\epsilon) = L - \frac{\partial L}{\partial A_{\mu,\sigma}} A_{\mu,\rho} N_\sigma N^\rho - \frac{\partial L}{\partial \phi_{,\rho}} \phi_{,\sigma} N^\sigma N_\rho - \frac{\partial L}{\partial \phi_{,\rho}^*} \phi_{,\sigma}^* N^\sigma N_\rho, \quad (4.7)$$

and expressing $T(\epsilon)$ as a function of A_μ , ϕ , ϕ^* , B^μ , χ , χ^* and the space derivatives of A_μ , ϕ and ϕ^* , we get (in a natural co-ordinate system)

$$\begin{aligned} T(\epsilon) &= [\frac{1}{4}F_{ij}F^{ij} - \frac{1}{2}g^{ij}B_iB_j - \frac{1}{2}g^{44}B_4B_4 - \chi\chi^* \\ &\quad - \frac{1}{2}\sqrt{(g^{44})}\{B_iA_{4,i} + A_{4,i}B_i\}g^{ij} + \frac{1}{2}\sqrt{(g^{44})}\{B_4A_{,i}^i + A_{,i}^iB_4\} + \phi_{,i}^*\phi_{,j}g^{ij}] \\ &\quad + \epsilon[\phi^*A_4\chi^*\sqrt{(g^{44})} - \chi A_4\phi\sqrt{(g^{44})} + g^{ij}(A_i\phi^*\phi_{,j} - \phi_{,i}^*\phi A_j)] - \epsilon^2[A_i\phi^*\phi A_jg^{ij}]. \end{aligned} \quad (4.8)$$

Therefore the interaction Hamiltonian $T_{\text{int.}}$ is given by

$$T_{\text{int.}} = T(\epsilon) - T(0) \\ = \epsilon[\phi^* A_4 \chi^* \sqrt{(g^{44})} - \chi A_4 \phi \sqrt{(g^{44})} + g^{ij}(A_i \phi^* \phi_{,j} - \phi_{,i}^* \phi A_j)] - \epsilon^2[A_i \phi^* \phi A_j g^{ij}]. \quad (4.9)$$

Substituting in (4.9) for χ and χ^* from (4.6) this becomes

$$T_{\text{int.}} = \epsilon[g^{44}\{\phi^* A_4 \phi_{,4} - \phi_{,4}^* A_4 \phi\} + g^{ij}\{\phi^* A_i \phi_{,j} - \phi_{,j}^* A_i \phi\}] \\ - \epsilon^2 \phi^* \phi [A_i A_j g^{ij} + 2A_4 A_4 g^{44}]. \quad (4.10)$$

$T_{\text{int.}}$ is an invariant which in a natural co-ordinate system takes the form (4.10). Clearly the expression for $T_{\text{int.}}$ which is valid in any co-ordinate system (and which reduces to (4.10) in a natural co-ordinate system) is

$$T_{\text{int.}} = \epsilon g^{\sigma\tau}[\phi^* A_\sigma \phi_{,\tau} - \phi_{,\tau}^* A_\sigma \phi] - \epsilon^2 \phi^* \phi [A_\sigma A_\tau g^{\sigma\tau} + N^\sigma N^\tau A_\sigma A_\tau]. \quad (4.11)$$

This is the interaction Hamiltonian found by Kanesawa & Tomonaga.

REFERENCES

- Belinfante, F. J. 1949 *Phys. Rev.* **76**, 66.
 Dirac, P. A. M. 1947 *The principles of quantum mechanics*, 3rd ed. Oxford: Clarendon Press.
 Kanesawa, S. & Tomonaga, S. 1948a *Progr. Theor. Phys.* **3**, 1.
 Kanesawa, S. & Tomonaga, S. 1948b *Progr. Theor. Phys.* **3**, 101.
 Miyomoto, Y. 1948 *Progr. Theor. Phys.* **3**, 124.
 Schwinger, J. 1948 *Phys. Rev.* **74**, 1439.
 Tomonaga, S. 1946 *Progr. Theor. Phys.* **1**, 27.
 Veblen, O. 1933 *Invariants of differential quadratic forms*. Cambridge University Press.
 Wentzel, G. 1943 *Einführung in die Quantentheorie der Wellenfelder*. Vienna: Franz Deuticke.
 de Wet, J. S. 1948 *Proc. Roy. Soc. A*, **195**, 365.
 de Wet, J. S. 1950 *Proc. Camb. Phil. Soc.* (in the press).

The forces on a submerged spheroid moving in a circular path

BY T. H. HAVELOCK, F.R.S.

(Received 28 December 1949)

Expressions are obtained for the tangential and radial forces on a sphere moving in a circular path at constant depth; similar calculations are made for a prolate spheroid, including in this case the couple acting on the spheroid. Numerical computations have been made, and curves are given to show the effect of curvature of the path upon the wave resistance.

1. The forces on a ship moving in a curved path are, no doubt, affected to some extent by the wave motion produced, but it is not easy to estimate the magnitude or nature of this influence. In the following paper an approach is made to some aspects of this problem by considering some cases of a submerged body moving in a circular path, namely, a sphere and a prolate spheroid. The motion of a sphere has been examined recently by Sretensky (1946), but the results given by him are incorrect. In the present work a different method is adopted; it is one which can be used for bodies of other forms, and also for non-uniform motion.

2. We may derive first expressions for the ideal case of a simple source moving in any manner at constant depth f below the free surface of the water. We take fixed axes with O in the free surface, Oz vertically upwards, and we use cylindrical coordinates (ϖ, θ, z) . If at time τ the strength of the source is m and its horizontal distance from O is ϖ_0 , the velocity potential due to an infinitesimal step in the motion is given, as in equation (27) of a previous paper (Havelock 1949), by

$$\phi = 2mg^{\frac{1}{2}}\delta\tau \int_0^\infty J_0(\kappa\varpi_0) e^{-\kappa(f-z)} \sin\{g^{\frac{1}{2}}\kappa^{\frac{1}{2}}(t-\tau)\} \kappa^{\frac{1}{2}} d\kappa. \quad (1)$$

We may regard the effect due to a point source, varying in strength and moving in any manner, as made up of the superposition of small steps of this nature. In particular, for the present problem, we suppose the source of constant strength and to be moving in a circle of radius h ; further, we take the motion to start at $t = 0$ and the angular velocity to have a constant value Ω . Hence we obtain the velocity potential at any time t as

$$\phi = \frac{m}{r_1} - \frac{m}{r_2} + 2mg^{\frac{1}{2}} \int_0^t d\tau \int_0^\infty J_0(\kappa\varpi_0) e^{-\kappa(f-z)} \sin\{g^{\frac{1}{2}}\kappa^{\frac{1}{2}}(t-\tau)\} \kappa^{\frac{1}{2}} d\kappa, \quad (2)$$

where

$$\left. \begin{aligned} r_1^2 &= \varpi^2 + h^2 - 2\varpi h \cos(\theta - \Omega t) + (z + f)^2, \\ r_2^2 &= \varpi^2 + h^2 - 2\varpi h \cos(\theta - \Omega t) + (z - f)^2, \\ \varpi_0^2 &= \varpi^2 + h^2 - 2\varpi h \cos(\theta - \Omega\tau). \end{aligned} \right\} \quad (3)$$

For the relative steady state which is ultimately established we require the limiting form of (2) as $t \rightarrow \infty$. We substitute in (2)

$$J_0(\kappa\varpi_0) = J_0(\kappa\varpi) J_0(\kappa h) + 2 \sum_{n=1}^\infty J_n(\kappa\varpi) J_n(\kappa h) \cos n(\theta - \Omega\tau). \quad (4)$$

We then integrate with respect to τ term by term and obtain the limiting form of the resulting integrals in κ as $t \rightarrow \infty$. This process readily gives the result

$$\phi = \frac{m}{r_1} - \frac{m}{r_2} + 4m \sum_1^\infty P \int_0^\infty \frac{\kappa e^{-\kappa(f-z)}}{\kappa - n^2 \Omega^2/g} J_n(\kappa \varpi) J_n(\kappa h) \cos n(\theta - \Omega t) \\ - \frac{4\pi m \Omega^2}{g} \sum_1^\infty n^2 J_n(n^2 \Omega^2 \varpi/g) J_n(n^2 \Omega^2 h/g) \exp[-n^2 \Omega^2(f-z)/g] \sin n(\theta - \Omega t), \quad (5)$$

where P denotes the principal value of the integral. It may be verified directly that this solution satisfies the conditions for the quasi-steady state.

3. If a sphere, of radius a , is moving uniformly in a circle we may, as a first approximation, take it as equivalent to a doublet of moment M equal to $\frac{1}{2}a^3 h \Omega$. It is easily seen that the velocity potential for this doublet can be derived from (5) by taking $\partial\phi/\partial t$ and replacing $mh\Omega$ by M . Thus we obtain

$$\phi = \frac{M \varpi \sin(\theta - \Omega t)}{r_1^3} - \frac{M \varpi \sin(\theta - \Omega t)}{r_2^3} \\ + \frac{4M}{h} \sum_1^\infty P \int_0^\infty \frac{\kappa e^{-\kappa(f-z)}}{\kappa - n^2 \Omega^2/g} n J_n(\kappa \varpi) J_n(\kappa h) \sin n(\theta - \Omega t) d\kappa \\ + \frac{4\pi M \Omega^2}{gh} \sum_1^\infty n^3 J_n(n^2 \Omega^2 \varpi/g) J_n(n^2 \Omega^2 h/g) \exp[-n^2 \Omega^2(f-z)/g] \cos n(\theta - \Omega t). \quad (6)$$

It may be noted that the second term in (6) is equivalent to

$$\frac{2M \varpi \sin(\theta - \Omega t)}{r_2^3} + \frac{4M \Omega^2}{gh} \sum_1^\infty P \int_0^\infty \frac{e^{-\kappa(f-z)}}{\kappa - n^2 \Omega^2/g} n^3 J_n(\kappa \varpi) J_n(\kappa h) \sin n(\theta - \Omega t) d\kappa. \quad (7)$$

We may deduce the wave resistance from the energy propagated outwards through a cylindrical surface, namely,

$$- \int_{-\infty}^0 dz \int_0^{2\pi} \rho \frac{\partial \phi}{\partial t} \frac{\partial \phi}{\partial \varpi} \varpi d\theta. \quad (8)$$

Taking the cylinder of large radius, we require the first terms in the expansion of (6), which are seen to be of order $\varpi^{-\frac{1}{2}}$. One such term comes from the integral in (7). Referring to (4), since we are concerned with large values of ϖ , we may replace $J_n(\kappa \varpi)$ in the expansion by $H_n^{(1)}(\kappa \varpi)$ and take the real part. Thus we have to evaluate the real part of

$$P \int_0^\infty \frac{e^{-\kappa(f-z)}}{\kappa - \kappa_n} H_n^{(1)}(\kappa \varpi) J_n(\kappa h) d\kappa \quad (\varpi > h; z < 0), \quad (9)$$

where we have put $\kappa_n = n^2 \Omega^2/g$. Regarding κ as a complex variable, we may change the path of integration to the positive half of the imaginary axis; taking account of the indentation at $\kappa = \kappa_n$, we obtain for (9)

$$\frac{2}{\pi} \int_0^\infty \frac{m \sin m(f-z) + \kappa_n \cos m(f-z)}{m^2 + \kappa_n^2} K_n(\varpi m) I_n(hm) dm \\ - \pi Y_n(\kappa_n \varpi) J_n(\kappa_n h) \exp[-\kappa_n(f-z)]. \quad (10)$$

Collecting the results from (6), (7) and (10), and using the asymptotic expansions for J_n and Y_n , we obtain, for ϖ large,

$$\phi \sim \frac{4M \Omega}{h} \left(\frac{2\pi'}{g\varpi} \right)^{\frac{1}{2}} \sum_1^\infty n^2 J_n(\kappa_n h) \exp[-\kappa_n(f-z)] \cos \{n(\theta - \Omega t - \frac{1}{2}\pi) + \kappa_n \varpi - \frac{1}{4}\pi\}. \quad (11)$$

With this value of ϕ in (8) we obtain the rate of propagation of energy outwards, and this must be equal to $Rh\Omega$ with R the steady wave resistance. Finally, replacing M by $\frac{1}{2}a^3h$, we obtain for the wave resistance of the sphere

$$R = \frac{4\pi^2\rho a^6\Omega^4}{gh} \sum_1^\infty n^5 J_n^2(n^2\Omega^2h/g) \exp[-2n^2\Omega^2f/g]. \quad (12)$$

It is of interest to examine the limiting form of this expression as $h \rightarrow \infty$, $\Omega \rightarrow 0$, $h\Omega \rightarrow c$. The series then becomes an integral, and by using appropriate asymptotic expansions for Bessel functions of large order and large argument, it is found that (12) reduces to

$$R = 4\pi\rho a^6\kappa_0^4 c^2 \int_0^{\frac{1}{2}\pi} \sec^5\beta \exp[-2\kappa_0 f \sec^2\beta] d\beta, \quad (13)$$

with $\kappa_0 = g/c^2$, and this is the wave resistance for a sphere in steady rectilinear motion with velocity c .

4. Returning to the general expression (6) we may evaluate the resultant fluid pressure on the sphere and so obtain both the radial force and the tangential force or wave resistance. The effective part of the pressure comes from $\rho\partial\phi/\partial t$, and we notice from (6) that the terms divide into two groups, (i) those symmetrical in the angle $\theta - \Omega t$, (ii) those anti-symmetrical in that angle. Obviously the resultant radial force on the sphere comes from the terms in group (i), while the tangential force is due to those in group (ii). It is necessary to note that, in using this method, the expression for the velocity potential must be carried to a further degree of approximation, because the boundary condition at the surface of the sphere must be satisfied to the same stage. Let (r, α, β) be spherical polar co-ordinates referred to the centre of the sphere so that

$$\left. \begin{aligned} \varpi \cos(\theta - \Omega t) &= h + r \sin \alpha \cos \beta, \\ \varpi \sin(\theta - \Omega t) &= r \sin \alpha \sin \beta, \\ z &= -f + r \cos \alpha. \end{aligned} \right\} \quad (14)$$

The first term in (6) is the doublet D giving the correct normal velocity at the surface of the sphere. The remaining terms in (6) may be expanded in the neighbourhood of the sphere in spherical harmonics so that we have (6) in the form

$$\phi = D + \sum_1^\infty (r/a)^n S_n(\alpha, \beta). \quad (15)$$

The required extension is then

$$\phi = D + \sum_1^\infty \left\{ \left(\frac{r}{a} \right)^n + \frac{n}{n+1} \left(\frac{a}{r} \right)^{n+1} \right\} S_n(\alpha, \beta). \quad (16)$$

Taking the tangential resultant force, the effective terms in $\rho\partial\phi/\partial t$ from (6) are

$$(2\pi\rho a^3\Omega^2/g) \sum_1^\infty n^4 J_n(\kappa_n \varpi) J_n(\kappa_n h) \exp[-\kappa_n(f-z)] \sin n(\theta - \Omega t). \quad (17)$$

In this, we put

$$\begin{aligned} & J_n(\kappa_n \varpi) \exp[-\kappa_n(f-z)] \sin n(\theta - \Omega t) \\ &= \frac{(-i)^n}{2\pi} \exp[-\kappa_n(f-z)] \int_{-\pi}^{\pi} \exp[i\kappa_n \varpi \cos(\theta - \Omega t - u)] \sin nu du \\ &= \frac{(-i)^n}{2\pi} \exp[-2\kappa_n f] \int_{-\pi}^{\pi} \exp[i\kappa_n r(\sin \alpha \cos \beta \cos u + \sin \alpha \sin \beta \sin u - i \cos \alpha) \\ &\quad - i\kappa_n h \cos u] \sin nu du. \end{aligned} \quad (18)$$

Expanding under the integral sign, we obtain the required expression in terms of spherical harmonics. Obtaining the resultant force involves multiplying the pressure by a surface harmonic of the first order and integrating over the sphere. Thus we only need the first-order term from (18), which is

$$\frac{(-i)^n}{2\pi} i\kappa_n a \exp[-2\kappa_n f] \int_{-\pi}^{\pi} (\sin \alpha \cos \beta \cos u + \sin \alpha \sin \beta \sin u - i \cos \alpha) \exp[-i\kappa_n h \cos u] \sin n u du. \quad (19)$$

In accordance with (16), this must be multiplied by $\frac{3}{2}$ to get the correct operative value of the pressure. We insert these results in (17), multiply by $\sin \alpha \sin \beta$ and integrate over the surface of the sphere. It is easily verified that this process gives the same expression (12) for the tangential resistance.

5. For the resultant radial force outwards, we carry out the same process on the pressure derived from the first three terms of (6), noting that in accordance with (16), the first-order surface harmonic from the second and third terms of (6) must be multiplied by $\frac{3}{2}$. We then multiply by $\sin \alpha \cos \beta$ and integrate over the surface of the sphere. The details of the calculation need not be given; after some reduction, we find for the resultant radial force outwards the expression

$$\frac{2}{3}\pi\rho a^3 h \Omega^2 \left(1 - \frac{3}{16} \frac{a^3}{f^3}\right) + 4\pi\rho a^6 \Omega^2 \sum_1^{\infty} P \int_0^{\infty} \frac{\kappa^2 e^{-2\kappa f}}{\kappa - n^2 \Omega^2 / g} n^2 J_n(\kappa h) J'_n(\kappa h) d\kappa. \quad (20)$$

This expression does not lend itself readily to numerical computation. We notice, however, that the first term in (20) represents an effective mass $\frac{2}{3}\pi\rho a^3(1 - 3a^3/16f^3)$, which is the first approximation for a sphere under a free surface, neglecting gravity. On the other hand, when the angular velocity is small the last term in (20) approximates to $\frac{1}{4}\pi\rho a^6 \Omega^2 / f^3$, since $\sum n^2 J_n(\kappa h) J'_n(\kappa h) = \frac{1}{4}\kappa h$. Thus for small velocity, the effective mass approximates to $\frac{2}{3}\pi\rho a^3(1 + 3a^3/16f^3)$, as for a sphere under a rigid surface.

It is of some interest to make calculations from (12), so as to obtain some idea of the nature and magnitude of the effect of curvature of the path upon the wave resistance. Curves showing the results are given in figure 1. The abscissae are values of $h\Omega/\sqrt{gf}$, so as to include rectilinear motion for comparison; the ordinates are values of $R/M'g(a/f)^3$, where M' is the mass displaced by the sphere. Curve *A* is for steady rectilinear motion, that is, for the limiting case $h/f \rightarrow \infty$, and calculated from (13). Curve *B* is for $h = f$. Even in this case the mean curve approximates to *A*, but it is of interest to note the hump and hollows due to wave interference when the sphere is making complete circles. For curve *C* we have taken $h = 4f$; it shows how with increasing radius of the circular path these interference effects disappear and the wave resistance approximates quite closely to that for straight-line motion at the same linear speed.

6. We consider now a prolate spheroid with its axis at a constant depth f below the surface, its centre C describing a horizontal circle of radius h with constant angular velocity Ω , the axis of the spheroid remaining at right angles to the rotating radius through C . We use the same fixed axes as before, with cylindrical co-ordinates $O(\varpi, \theta, z)$; and, when required, we use rotating axes $C(x, y, z)$ with Cx along the axis of the spheroid in the direction of motion and Cz vertically upwards.

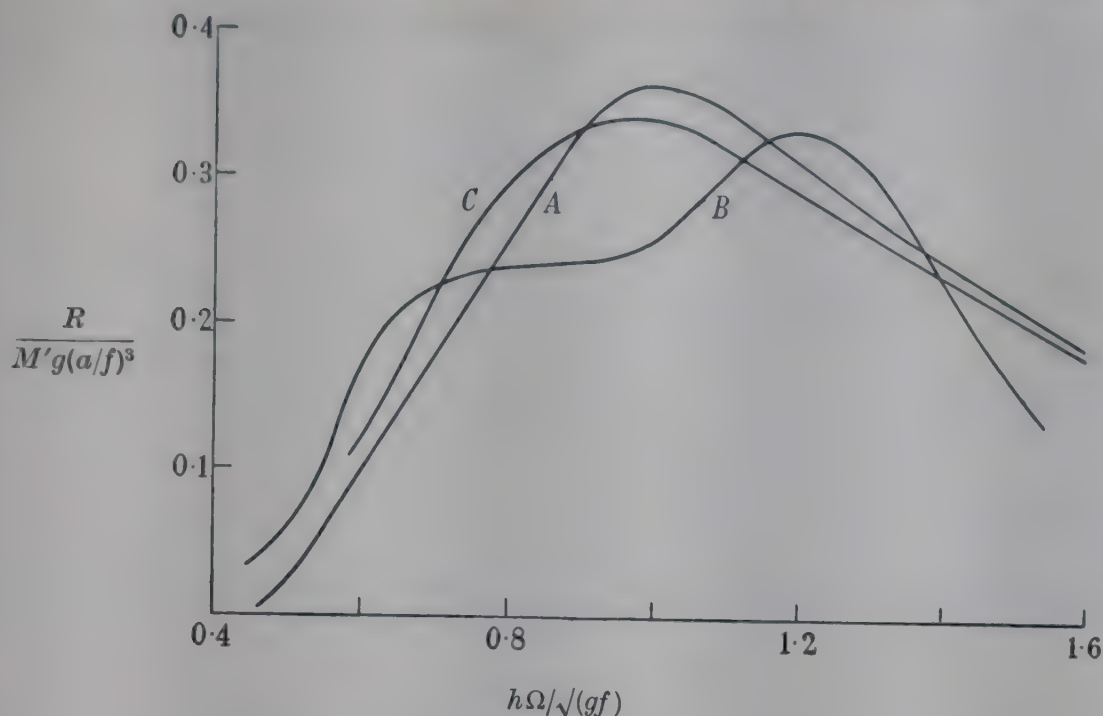


FIGURE 1

The motion of the spheroid is made up of a linear velocity $h\Omega$ parallel to Cx and a rotation Ω about Cz . In terms of spheroidal co-ordinates given by

$$x = ae\mu\zeta, \quad y = ae(1 - \mu^2)^{\frac{1}{2}}(\zeta^2 - 1)^{\frac{1}{2}}\cos\omega, \quad z = ae(1 - \mu^2)^{\frac{1}{2}}(\zeta^2 - 1)^{\frac{1}{2}}\sin\omega, \quad (21)$$

the known solution for this motion in an infinite liquid is (Lamb 1932)

$$\phi = 2Aae h\Omega AP_1(\mu)Q_1(\zeta) - \frac{2}{3}Ba^2e^2\Omega P_2^1(\mu)Q_2^1(\zeta)\cos\omega, \quad (22)$$

$$\text{with } \left. \begin{aligned} A^{-1} &= 2e/(1 - e^2) - \log\{(1 + e)/(1 - e)\}, \\ B^{-1} &= \{3(2 - e^2)/e^2\}\log\{(1 + e)/(1 - e)\} - 2(6 - 7e^2)/e(1 - e^2). \end{aligned} \right\} \quad (23)$$

It is well known that the linear motion can be expressed in terms of a certain source distribution along the axis of the spheroid, and it can easily be shown that the angular motion can be ascribed to a doublet distribution along the axis. In fact, (22) is equivalent to

$$\phi = Ah\Omega \int_{-ae}^{ae} \frac{k dk}{\{(x - k)^2 + y^2 + z^2\}^{\frac{3}{2}}} + B\Omega y \int_{-ae}^{ae} \frac{k(a^2e^2 - k^2) dk}{\{(x - k)^2 + y^2 + z^2\}^{\frac{5}{2}}}. \quad (24)$$

We may now obtain the required solution by integration of the expression for a source given in (6). For the first term in (24) we have to replace a typical factor $J_n(\kappa h)\cos n(\theta - \Omega t)$ in (6) by $J_n\{\kappa(h^2 + k^2)^{\frac{1}{2}}\}\cos n(\theta - \Omega t - \alpha)$, where $\tan\alpha = k/h$; and, taking account of the integration in k , this may be replaced by

$$J_n\{\kappa(h^2 + k^2)^{\frac{1}{2}}\}\sin n\alpha \sin n(\theta - \Omega t).$$

Further, so far as the co-ordinates x, y, z are concerned, the second term in (24) may be derived by taking $\partial/\partial y$ of the first term; and when the expressions are put in terms of the fixed co-ordinates this is equivalent to operating by $\partial/\partial h$. Also we have

$$\begin{aligned} &\frac{\partial}{\partial h} [J_n\{\kappa(h^2 + k^2)^{\frac{1}{2}}\}\sin n\alpha] \\ &= \frac{1}{2}\kappa [J_{n-1}\{\kappa(h^2 + k^2)^{\frac{1}{2}}\}\sin(n-1)\alpha - J_{n+1}\{\kappa(h^2 + k^2)^{\frac{1}{2}}\}\sin(n+1)\alpha]. \end{aligned} \quad (25)$$

Carrying out these operations we obtain for the rotating spheroid at depth f , its centre describing a circle of radius h ,

$$\phi = \int_{-ae}^{ae} \{Ah\Omega kF + B\Omega k(\alpha^2 e^2 - k^2)G\} dk, \quad (26)$$

with

$$F = \frac{1}{r_1} - \frac{1}{r_2} + \sum_1^\infty P \int_0^\infty \frac{\kappa e^{-\kappa(f-z)}}{\kappa - \kappa_n} J_n(\kappa\varpi) J_n\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \sin n\alpha \sin n(\theta - \Omega t) d\kappa \\ + (4\pi\Omega^2/g) \sum_1^\infty n^2 J_n(\kappa_n\varpi) J_n\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \exp[-\kappa_n(f-z)] \sin n\alpha \cos n(\theta - \Omega t), \quad (27)$$

$$G = \frac{h - \varpi \cos(\theta - \Omega t)}{r_1^3} - \frac{h - \varpi \cos(\theta - \Omega t)}{r_2^3} \\ + \sum_1^\infty P \int_0^\infty \frac{\kappa^2 e^{-\kappa(f-z)}}{\kappa - \kappa_n} J_n(\kappa\varpi) [J_{n+1}\{\kappa(h^2 + k^2)^{\frac{1}{2}}\} \sin(n+1)\alpha \\ - J_{n-1}\{\kappa(h^2 + k^2)^{\frac{1}{2}}\} \sin(n-1)\alpha] \sin n(\theta - \Omega t) d\kappa \\ + \frac{2\pi\Omega^4}{g^2} \sum_1^\infty n^4 J_n(\kappa_n\varpi) [J_{n+1}\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \sin(n+1)\alpha \\ - J_{n-1}\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \sin(n-1)\alpha] \exp[-\kappa_n(f-z)] \cos n(\theta - \Omega t). \quad (28)$$

In this $\kappa_n = n^2\Omega^2/g$, $\tan \alpha = k/h$, and

$$\left. \begin{aligned} r_1^2 &= \{\varpi \sin(\theta - \Omega t) - k\}^2 + \{h - \varpi \cos(\theta - \Omega t)\}^2 + (z+f)^2, \\ r_2^2 &= \{\varpi \sin(\theta - \Omega t) - k\}^2 + \{h - \varpi \cos(\theta - \Omega t)\}^2 + (z-f)^2. \end{aligned} \right\} \quad (29)$$

By comparison with § 3, we see that for ϖ large we have

$$\phi \sim 2h\Omega^2 \left(\frac{2\pi}{g\varpi}\right)^{\frac{1}{2}} \sum_1^\infty \{2nAL_n + n^3B(\Omega^2/gh)M_n\} \\ \exp[-\kappa_n(f-z)] \cos\{n(\theta - \Omega t - \frac{1}{2}\pi) + \kappa_n\varpi - \frac{1}{4}\pi\}, \quad (30)$$

$$\left. \begin{aligned} \text{with } L_n &= \int_{-ae}^{ae} k J_n\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \sin n\alpha dk, \\ M_n &= \int_{-ae}^{ae} k(\alpha^2 e^2 - k^2) [J_{n+1}\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \sin(n+1)\alpha \\ &\quad - J_{n-1}\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \sin(n-1)\alpha] dk. \end{aligned} \right\} \quad (31)$$

Using this in (8) we obtain the rate of propagation of energy outwards; if R is the tangential resistance and G the couple required to maintain the uniform motion, this leads to

$$Rh\Omega + G\Omega = \frac{4\pi^2\rho h^2\Omega^5}{g} \sum_1^\infty n^3 \left(2AL_n + B\frac{\Omega^2}{gh}n^2M_n\right)^2 \exp[-2n^2\Omega^2f/g]. \quad (32)$$

7. We may obtain the resistance, the couple and the radial force by calculating the resultant fluid pressure on the spheroid. For the wave resistance the only part of the pressure which gives a resultant is the term $\rho\partial\phi/\partial t$, and we have

$$R = \iint p l dS = \rho a^2(1-e^2) \int_{-1}^1 \int_0^{2\pi} \frac{\partial\phi}{\partial t} \mu d\mu d\omega, \quad (33)$$

in terms of the co-ordinates (21), with $\zeta = \zeta_0 = 1/e$ on the spheroid. Taking from (26) the terms which contribute to the resultant tangential force, we have

$$\frac{\partial \phi}{\partial t} = \int_{-ae}^{ae} \{A h \Omega k F_1 + B \Omega k (a^2 e^2 - k^2) G_1\} dk, \quad (34)$$

with

$$F_1 = \frac{4\pi\Omega^3}{g} \sum_1^\infty n^3 J_n(\kappa_n \varpi) J_n\{\kappa_n(h^2 + k^2)^{\frac{1}{2}}\} \exp[-\kappa_n(f-z)] \sin n\alpha \sin n(\theta - \Omega t), \quad (35)$$

and a similar expression for G_1 , derived from (28). If (34) is expanded in the form

$$\frac{\partial \phi}{\partial t} = \sum_{r=1}^\infty \sum_{s=0}^r (A_r^s \cos s\omega + B_r^s \sin s\omega) P_r^s(\mu) P_r^s(\zeta), \quad (36)$$

we must add a similar expression with $Q_r^s(\zeta)$ in place of $P_r^s(\zeta)$ so as to maintain the boundary condition at the surface of the spheroid; for this part of ϕ this is $\partial\phi/\partial\zeta = 0$ for $\zeta = \zeta_0$. Hence on the spheroid we have

$$\frac{\partial \phi}{\partial t} = \sum_{r=1}^\infty \sum_{s=0}^r C_r^s (A_r^s \cos s\omega + B_r^s \sin s\omega) P_r^s(\zeta_0) P_r^s(\mu),$$

with
$$C_r^s = 1 - P_r^{s'}(\zeta_0) Q_r^s(\zeta_0) / P_r^s(\zeta_0) Q_r^{s'}(\zeta_0). \quad (37)$$

We now expand (34) in the form (36), noting that for the value of R we only require the term in $P_1^0(\mu)$. For this purpose we have

$$\begin{aligned} & J_n(\kappa_n \varpi) \exp[\kappa_n(f-z)] \sin n(\theta - \Omega t) \\ &= \frac{(-i)^n}{2\pi} \exp[-2\kappa_n f + \kappa_n z] \int_{-\pi}^{\pi} \exp[i\kappa_n(x \sin u - y \cos u) + i\kappa_n h \cos u] \sin nu du, \end{aligned} \quad (38)$$

with the origin now at the centre of the spheroid. Substituting from (21), we multiply by $\mu d\mu d\omega$ and integrate over the surface of the spheroid. It can be shown that

$$\begin{aligned} & \int_{-1}^1 \mu d\mu \int_0^{2\pi} \exp[i\kappa_n\{a\mu \sin u - b(1-\mu^2)^{\frac{1}{2}}(\cos u \cos \omega + i \sin \omega)\}] d\omega \\ &= 4\pi i (\pi/2\kappa_n a e^3)^{\frac{1}{2}} \sin^{-\frac{1}{2}} u J_{\frac{1}{2}}(\kappa_n a e \sin u). \end{aligned} \quad (39)$$

Using (39) in (38), we have, so far as this typical term is concerned, the integral

$$4\pi i (\pi/2\kappa_n a e^3)^{\frac{1}{2}} \int_{-\pi}^{\pi} \exp[i\kappa_n h \cos u] J_{\frac{1}{2}}(\kappa_n a e \sin u) \sin^{-\frac{1}{2}} u \sin nu du. \quad (40)$$

It is of interest to find that (40) can be put in the form $(4\pi^2/a^2 e^3)^{\frac{1}{2}} i^n L_n$, in the notation of (31). Collecting these results and including the factor C_1^0 from (37), we obtain for the wave resistance

$$R = \frac{8\pi^2 \rho h \Omega^4}{g} A \sum_1^\infty n^3 L_n \left(2A L_n + B \frac{\Omega^2}{gh} n^2 M_n \right) \exp[-2n^2 \Omega^2 f/g]. \quad (41)$$

We could obtain the couple G by similar calculations; or, using (32), we have

$$G = \frac{4\pi^2 \rho h \Omega^6}{g^2} B \sum_1^\infty n^5 M_n \left(2A L_n + B \frac{\Omega^2}{gh} n^2 M_n \right) \exp[-2n^2 \Omega^2 f/g]. \quad (42)$$

The radial force can be obtained by the same method from the remaining terms in the velocity potential; but the expressions are lengthy and not suitable for numerical computation.

If we take limiting values as $h \rightarrow \infty$, $\Omega \rightarrow 0$, $h\Omega \rightarrow c$, the couple G becomes zero and R reduces to the wave resistance given previously (Havelock 1931) for the linear motion of a spheroid. On the other hand, if we take $h = 0$, we find that $L_n = 0$, $M_{2n+1} = 0$, and we obtain the couple for pure rotation as

$$G = \frac{512\pi^2\rho a^6 e^6}{g} \Omega^4 B^2 \sum_1^\infty n^5 P_{2n}^2 \exp[-8n^2\Omega^2 f/g],$$

with
$$P_{2n} = \int_0^1 (1-u^2) J_{2n}(4n^2\Omega^2 aeu/g) du. \quad (43)$$

8. For numerical computation the integrals for L , M can be expanded in various forms; for instance, one which proved useful can be derived from the expansion

$$J_n\{p(1+u^2)^{\frac{1}{2}}\} \sin(n \tan^{-1} u) = (J_{n-1} + J_{n+1}) \left(\frac{1}{2}pu\right) - (J_{n-3} + 3J_{n-1} + 3J_{n+1} + J_{n+3}) \frac{(\frac{1}{2}pu)^3}{3!} + \dots, \quad (44)$$

the Bessel functions having the argument p . For some values of the parameters it was found more convenient to evaluate the integrals by direct quadrature.

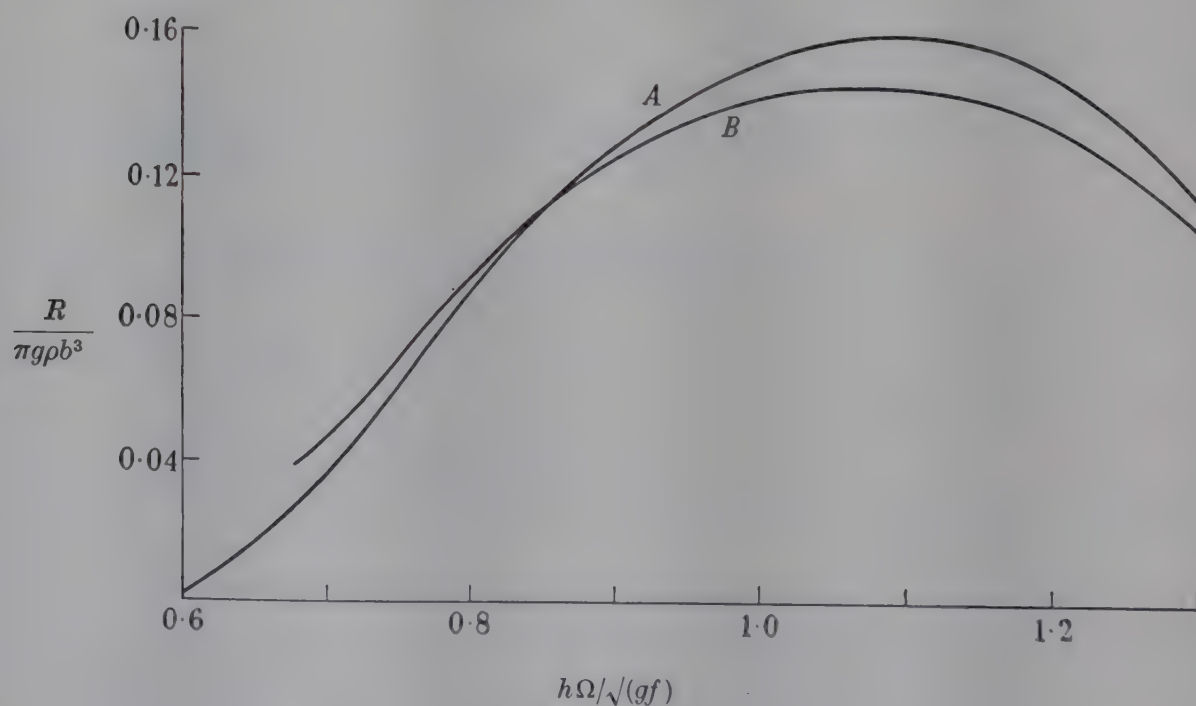


FIGURE 2

As a particular case we take a spheroid for which $2a = 5b$, so that $e = 0.9165$; and for the depth we take $f = 2b$. This was one of the cases for which calculations were made previously for rectilinear motion. To bring out the effect of curvature we take for the radius of the path $h = 5b$. The results for the wave resistance are shown in figure 2. The ordinates are values of $R/\pi g \rho b^3$, and the abscissae are $h\Omega/\sqrt{(gf)}$. The curve A is for linear motion and is taken from the paper already quoted (Havelock

1931). Curve *B* shows the effect of circular motion in this particular case; the difference between the two curves is quite small even in this rather extreme case. A curve is not given for the couple *G*, as the quantities *M* are rather difficult to evaluate with sufficient accuracy for this purpose; however, approximate computations were made and the maximum value appears to be at about $h\Omega/\sqrt{g\ell} = 1.25$ with a value of $G/\pi g \rho b^4$ of about 0.026. It might be expected that the couple would be small for a solid of revolution in this particular case; it would probably be larger for a flat ellipsoid, for which similar calculations could be made by the methods used in the present work, the appropriate source and doublet distributions being then over the plane area enclosed by the elliptic focal conic.

REFERENCES

- Havelock, T. H. 1931 *Proc. Roy. Soc. A*, **131**, 275.
 Havelock, T. H. 1949 *Quart. J. Mech. Appl. Math.* **2**, 332.
 Lamb, H. 1932 *Hydrodynamics*, 6th ed., p. 141. Cambridge University Press.
 Sretensky, L. 1946 *C.R. Acad. Sci. U.R.S.S.* **44**, no. 9.

The thermal accommodation coefficients of gases. I. An investigation of the effect of flashing

BY J. G. M. BREMNER, *Jesus College, Oxford**

(Communicated by D. L. Chapman, F.R.S.—Received 9 July 1949)

Modifications to the hot-wire method of Knudsen (1911) for determining accommodation coefficients have been introduced by various workers in recent years. These methods have in common the flashing of the wire in the gas (Roberts 1930; Mann 1934) or *in vacuo* (Mann & Newell 1937) immediately before heat-loss measurements are begun. This treatment was originally suggested by Roberts (1930) and was assumed to give a 'clean' surface by removing adsorbed gas films. The heat loss from the wire was reduced by flashing, and this was ascribed to a reduction in the accommodation coefficient.

The fall in heat loss after flashing has been confirmed, but it is accounted for, not by a reduction in the accommodation coefficient, but by the persistence of the thermal effects of flashing. Thus the effect of flashing is accentuated by working with silvered and platinized vessels; it is reproduced by flashing an adjacent wire and is particularly evident when working *in vacuo*.

Abnormal temperature rises can also be observed when pipette methods are used to introduce the gas to an evacuated vessel (Mann & Newell 1937; Raines 1939; Rolf 1944). The results obtained in similar experiments were shown to be due mainly to the slowness of the diffusion process.

INTRODUCTION

Roberts (1930) found that a given current would raise the temperature of a metal filament surrounded by a gas at low pressures, to a higher temperature, after flashing, than before. This reduction in heat loss he ascribed solely to a reduction in the accommodation coefficient; impurities on the metal were assumed to give an initial high

* Now at Imperial Chemical Industries Ltd., Billingham-on-Tees, Co. Durham.

accommodation coefficient and their removal, by flashing, a clean surface with much lower accommodation coefficient. The heat loss from a wire which had been flashed did not remain constant but slowly increased to that observed before flashing. This increase in heat loss he ascribed to the slow redeposition of impurities on the wire surface, and an adjustment for this effect was made by extrapolating to the time when the flashing current was switched off. Such extrapolated values are much lower, e.g. 0.06 for helium on tungsten, than those previously published (Knudsen 1911*a, b*, 1930; Soddy & Berry 1910, 1911; Hughes & Bevan 1928; Chapman & Hall 1929). The flashing procedure of Roberts has been used by others with similar results (Michels 1932, 1933; Rowley & Evans 1935; Spivak & Zaitzev 1935; Spivak & Zacharjin 1936). Mann (1934) modified the procedure somewhat by taking heat-loss measurements before the wire had cooled, after flashing, to the region of the bath temperature. A further modification consisted in the flashing of the wire *in vacuo*, followed by the introduction of the gas with a pipette arrangement, and this also gave a transient reduction in the heat loss from the wire (Mann & Newell 1937; Rowley & Evans 1935; Raines 1939; Rolf 1944). Other workers (*inter alios* Amdur, Jones & Pearlman 1944; Thomas & Olmer 1943; Beeck 1936) do not use the flashing technique, and consequently widely differing values of the accommodation coefficient are to be found in the recent literature. Moreover, it seems clear (Johnson 1930; Farkas 1931; Langmuir & Blodgett 1932) that the effect of adsorbed gas is specific (cf. Hughes & Bevan 1928; Chapman & Hall 1929; Bonhoeffer & Rowley 1933), and is not always to increase the accommodation coefficient as Roberts assumed.

To aid in clarification, a critical examination of the effect of flashing on the heat loss from a metal surface has been made. The particular object of this work was to assess the reality of the assumptions made by Roberts and others as to their origin. Such assumptions have an important bearing on catalytic work.

EXPERIMENTAL

Method

The general method followed standard procedure in which heat-loss measurements were taken of a looped tungsten wire immersed in a gas at low pressure.

Apparatus

The helium was supplied as pure by the British Oxygen Co. The capillary seal on the containing bulb was broken by a magnetic hammer, comprising a soft iron cylinder sealed in glass, and the gas reduced to 5 cm. pressure by its transfer to a 2 l. bulb, *A*, figure 1. In this way, simple mercury cut-offs could be used in subsequent gas manipulation. Before use, the helium was purified by circulation through two charcoal traps, *B*, *C*, cooled in liquid air and protected from the deactivating influence of mercury by two traps also immersed in liquid air. Circulation was carried out by working a Töpler pump, *D*, of 1 l. capacity, which could also be used to pipette gas into the filament vessel.

The filament was taken from spools of thorium-free tungsten wire of nominal diameter 0.006 cm. obtained from Johnson, Matthey and Co. and the General

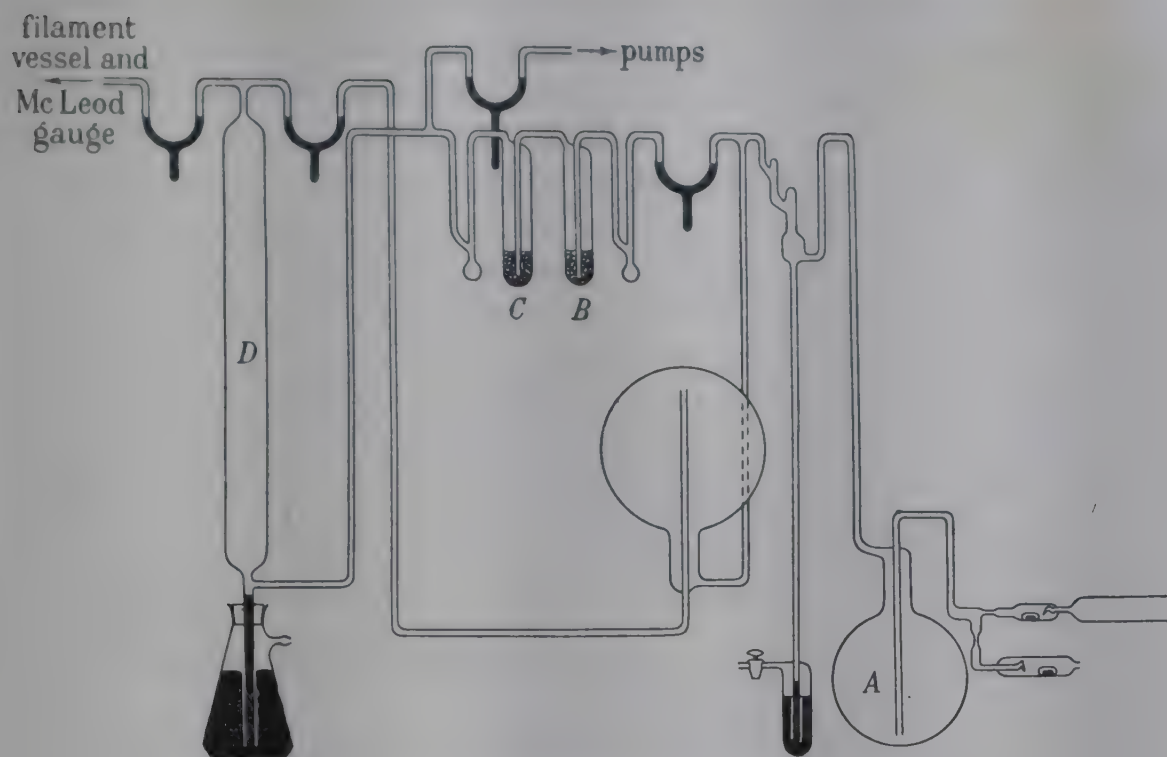


FIGURE 1. Helium storage and circulation system.
A, storage vessel; B, C, charcoal traps; D, Töpler pump.

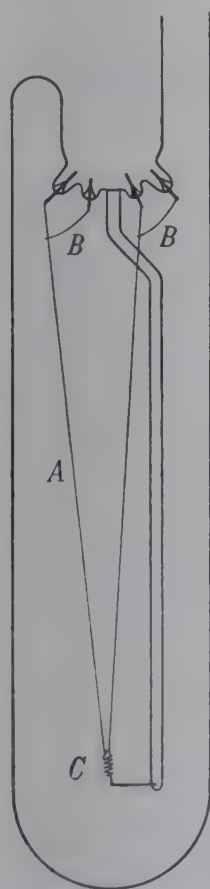


FIGURE 2. Arrangement of filament vessel.
A, tungsten filament; B, potential leads; C, molybdenum spring.

Electric Co. No difference was observed in their behaviour. The filament was mounted in a Pyrex vessel $30 \times 3\frac{1}{2}$ cm. which had four tungsten electrodes sealed into the head, *A*, figure 2. It was spot-welded to two of the symmetrically disposed electrodes, the remaining pair serving for potential leads, *B*. To eliminate junction effects, these leads were taken from the same spool as the filament. They were welded about 1.5 cm. from the electrodes. This difficult operation was carried out using machined copper electrodes worked in an atmosphere of sulphur-free hydrogen. The bottom loop of the filament was held by a molybdenum spring sealed into a pyrex rod, *C*. Connexion to the bridge circuit was made through copper leads, silver-soldered to nickel strips welded to the tungsten electrodes. After baking out, the filament vessel was immersed in a water thermostat whose temperature, normally 30° , was measured by a mercury-in-glass N.P.L. calibrated thermometer. Raising the bath temperature gave the temperature coefficient of resistance of the filament as $0.00413 \Omega \text{ deg.}^{-1}$ (General Electric Co.) and $0.00412 \Omega \text{ deg.}^{-1}$ (Johnson Matthey) between 30 and 55°C .

The filament vessel was protected from mercury vapour by two traps in series which were permanently immersed in liquid air.

A mercury seal of manometric height enabled the filament vessel to be removed, without admitting air to the main apparatus.

After assembly the entire apparatus was flamed, with the pumps running, to the point of collapse. The charcoal traps and filament vessel were baked out at 500°C for 6 to 10 hr. before use.

Pressures were measured on a McLeod gauge, 1 cm. displacement of which corresponded to a pressure of 10^{-5} cm. mercury.

Bridge circuit

The filament vessel formed one of the resistance arms of a Kelvin double bridge. All switches, including the galvanometer make-and-break, were of the mercury-cup type. The balancing resistance comprised a Cambridge millivoltmeter and shunt for current measurement and two resistance boxes (0 to 110 and 0 to 10,000 Ω) in parallel. The other pair of resistance arms were adjusted for equality by working the bridge to a null point both in its Kelvin and Wheatstone forms. The two ratios were 9.554 and 9.561, that is, within the accuracy of the resistance calibrations. The millivoltmeter and shunt were calibrated to give current readings on a Rayleigh type potentiometer. The current was maintained constant by balancing the e.m.f. developed across a constantan resistance against a Weston cell. The wire was flashed by connexion to the 105 V d.c. supply, and the temperature reached derived from its resistance using the data given by Jones (1926).

THE CALCULATION OF RESULTS

The thermal accommodation coefficient is the ratio of the heat removed from a surface to that which would be removed if the gas was raised to the surface temperature. The energy removed by the gas, ΔH_E , is obtained by subtracting that lost *in vacuo*, ΔH_v , from that removed at the same temperature in the presence of the gas, ΔH_g ,

$$\Delta H_E = \Delta H_g - \Delta H_v.$$

The heat, ΔH_c , which a monatomic gas at temperature T_1 of specific heat C_v , molecular weight M and at a pressure p (dynes cm.⁻²) may theoretically remove from a surface at a slightly higher temperature T_2 , is given, following Knudsen (1911), by

$$\Delta H_c = 7275p \frac{T_2 - T_1}{\sqrt{(T_1 M)}} \text{ ergs sec.}^{-1} \text{ cm.}^{-2}.$$

For most of this work $T_1 = 303^\circ \text{K}$ and, consequently, for helium striking a wire of length l cm. this equation, with p in cm. mercury, reduces to

$$\Delta H_c = 0.001255pl(T_2 - T_1) \text{ cal. sec.}^{-1}.$$

The thermal accommodation coefficient $a = (\Delta H_g - \Delta H_v)/\Delta H_c$. Values of ΔH_v , which increase almost linearly with rise in temperature, were obtained by heating the filament *in vacuo*. Accommodation coefficients, calculated after flashing, assume ΔH_c to have the value previously determined under steady conditions. It will be shown that this assumption is not justified under the conditions of our experiments. ΔH_c is temporarily decreased by flashing, and this is responsible for apparent low values of the accommodation coefficient. As ΔH_c depends directly on the gas pressure, p , the use of pipette methods for introducing known gas volumes will cause ΔH_c to increase with time until diffusion is complete. Accommodation coefficients calculated on the equilibrium pressure will consequently show an apparent increase with time. The slow rate of diffusion involved in the use of the pipette technique was demonstrated during the course of this work.

RESULTS

The results are given in two §§ 1 and 2. In the first the different procedures used by Roberts (1930), Mann (1934) and by Mann & Newell (1937) are examined. The second section is concerned with the elucidation of the origin of the effects of flashing.

(1) Techniques giving apparent low values of the calculated accommodation coefficient

(a) The Roberts procedure

A tungsten filament 38.5 cm. long was immersed in helium at 0.00485 cm. pressure. Using a very small current, about 0.002 A, the resistance of the wire at the bath temperature is found to be 7.630 Ω . The passage of 0.0225 A raised the wire temperature 18.84°C above that of the thermostat, corresponding to a resistance of 8.260 Ω —the 'preselected' resistance, AB , figure 3. Following Roberts's procedure the filament was raised to 2900° K for 1 min. and the resistance drop on cooling followed, using the small current, CD , figure 3. After the lapse of some $4\frac{1}{2}$ min. when the resistance of the filament had fallen to 7.80 Ω —much less than the preselected value—the larger predetermined current, 0.0225 A, was passed through the filament by operating a mercury-cup switch. The wire resistance increased rapidly, rose above the preselected resistance, reached a maximum at 8.58 Ω , and then slowly fell until the preselected resistance was once more recorded, in about 60 min., DFG , figure 3.

The value of the 'steady' accommodation coefficient calculated for the resistance rise 7.63 to 8.26 Ω was 0.16. A linear extrapolation of the resistance-time curve obtained after flashing to zero time gave 8.72 Ω as the resistance of a clean wire in

the Roberts sense. This would normally correspond to a wire temperature 34.6°C above that of the bath. Calculation gives the heat loss as $0.01055\text{ cal. sec.}^{-1}$, while that *in vacuo*, previously determined, is $0.00520\text{ cal. sec.}^{-1}$. Consequently the 'extrapolated' accommodation coefficient is

$$a = \frac{0.00535}{0.001255 \times 38.5 \times 0.00485 \times 34.6} = 0.066,$$

similar to that derived by Roberts (1930).

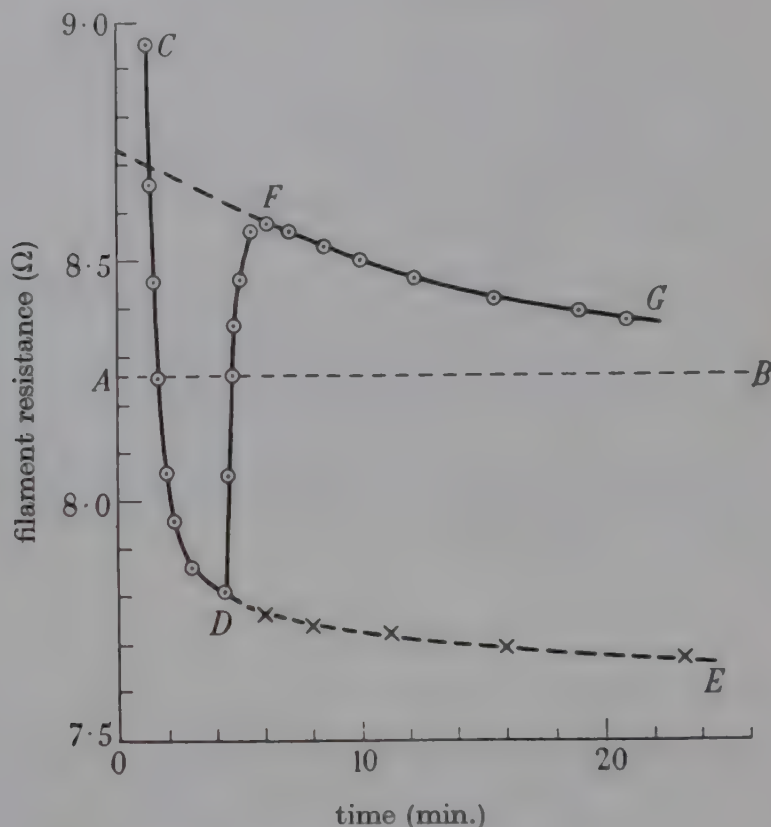


FIGURE 3. The Roberts procedure.

After many experiments of a similar type it was clear that the approach to the preselected resistance, after flashing, is not linear. Extrapolation along this curve would give a value for the derived accommodation coefficient much lower than 0.066. As figure 3 indicates, the normal cold resistance of the wire was not restored after flashing before the larger current of 0.0225 A was passed. A separate experiment, *DE*, figure 3, showed the time necessary for the return to the normal cold resistance to be about the same as that required for the higher resistance to return to its preselected value. Roberts states that with a small current passing, a 'steady' resistance is reached some 4 min. after flashing, and it was only then that the higher current, sensitive to a change in accommodation coefficient, is passed. In this experiment the larger current was passed when the resistance was only 2 % above its normal value and yet with this small increment we obtained a reduction of 50 % in the accommodation coefficient. Delaying the introduction of the higher current until the normal cold resistance was restored ($\sim 60\text{ min.}$) merely increased the resistance to the preselected value.

It may be mentioned that the steady accommodation coefficient 0.16 was obtained with a temperature increment of 19° , while that corresponding to the value 0.066 is 35°C . Experiment showed that the temperature coefficient of accommodation would not account for this difference. Flashing, moreover, does not cause a transient reduction in the gas concentration in the filament vessel as was shown by duplicating the effects of flashing with the vessel connected to only half the exterior volume used in the comparative experiment.

(b) *The Mann procedure*

In this experiment the predetermined current was passed when the wire resistance, after flashing, had fallen to the preselected value, AB , figure 4. The wire resistance did not remain at the preselected value but rose rapidly to a maximum and then slowly returned in about 60 min. to its former value. Linear extrapolation of the first few points on the falling curve, CD , gave 9.20Ω as the wire resistance at zero time corresponding to an accommodation coefficient of 0.035 for helium. This procedure, as Mann (1934) pointed out, gives even lower values for the accommodation coefficient than that used by Roberts (1930).

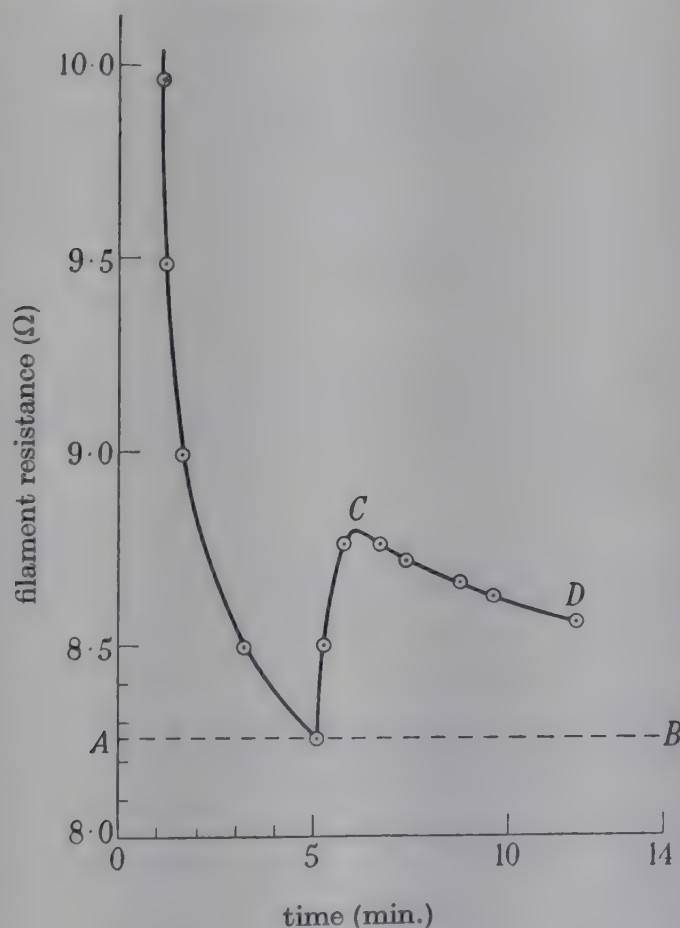


FIGURE 4. The Mann procedure.

(c) *The Mann-Newell procedure*

The tungsten filament was raised to 2900°C for 1 min. with the pumps running and, on cooling down, the gas (neon) was admitted to the filament vessel by using the Töpler pump as a pipette. 10 sec. after admitting the gas the current was

increased, when the resistance did not remain at its preselected value of 8.90Ω but rose rapidly to 9.1Ω and then began to fall slowly. Taking the time zero as that for the admission of the gas, the change in accommodation coefficient was calculated to be from 0.38, the steady value, to 0.20. Curvilinear extrapolation would give a value considerably lower than 0.20. That the assumption of progressive contamination, after flashing, was unjustified could easily be shown. On pumping out the neon and admitting a fresh charge of gas without flashing the wire, similar resistance changes were observed as those given when the wire had been flashed. The slow rate of diffusion was shown to be the cause of this phenomenon. The rate of passage of neon over a distance similar to that between the Töpler pump and the filament vessel was measured on the McLeod gauge. Diffusion was slow as it took over 60 sec. for the pressure to become steady. It was deduced, in this way, that the maximum rise in the wire resistance, after 40 sec. had elapsed, occurred when only 60 % of the equilibrium amount of gas was present in the filament vessel.

(2) *An examination of the thermal effects of flashing under various experimental conditions*

(a) *The effect produced by flashing using clear, silvered and platinized filament vessels*

In experiments 1 (a) and 1 (b) the filament vessels were not clear but had respectively a platinum and a silver film deposited on the internal walls of the vessel. The behaviour, on flashing, of such otherwise identical vessels is not the same. The results for experiments like 1 (a) and 1 (b) given by these vessels are compared, in table 1, with those by a clear vessel, after flashing at both 1300 and 2800 to 3000° K for times of from 1 to 10 and 15 min.

TABLE 1. VARIATION IN THE EXTRAPOLATED ACCOMMODATION COEFFICIENT WITH THE NATURE OF THE VESSEL WALL, THE FLASHING TEMPERATURE AND THE DURATION OF FLASHING

filament vessel	helium pressure (cm. Hg)	flashing temp. (° K)	flashing time (min.)	experiment type	resistance increment above the preselected (Ω)	extra-polated a.c.
silvered	—	3000	1	1 (b)	0.44	0.035
	0.00485	2900	1	1 (a)	0.28	0.07
	0.00466	3000	1	1 (b)	0.474	0.05
platinized	0.00466	3000	15	1 (b)	0.844	0.025
	0.00466	1300	1	1 (b)	0.044	0.145
	0.00466	1300	15	1 (b)	0.084	0.125
	0.00453	2800	1	1 (a)	0.004	0.16
	0.00453	2800	10	1 (a)	0.031	0.14
clear	0.00453	2800	1	1 (b)	0.005	0.14
	0.00453	2800	10	1 (b)	0.035	0.11

These results show:

- (i) the effects produced by flashing at 2800 to 3000° K were also observed at 1300° K but to a reduced extent;
- (ii) an increase of the flashing time increased the subsequent thermal effects;

(iii) the effects given by the clear vessel were much less than when the internal wall was silvered or platinized. An accommodation coefficient as low as 0.025 can be obtained by using the platinized vessel.

The hypotheses of Roberts (1930) and Mann (1934) must ascribe to these deposited films a marked effect in reducing the rate of contamination of the wire surface.

(b) *The effect of flashing on the accommodation coefficient derived at increasing wire temperatures*

Normally an accommodation coefficient is determined at a resistance corresponding to a wire temperature about 20° above that of the bath. As the resistance at which a measurement is made is increased the effect of flashing should also become more apparent so as to cause the same relative change in accommodation coefficient. The wire temperatures selected for this experiment were 3 , 15 and 45°C above that of the bath. The clear vessel was used, and after the wire had been flashed to 2800°K for 10 min. the appropriate current was passed at a time interval such as to give the maximum resistance reading 8 min. after flashing (see figure 5). Duplicate sets of experiments, *A* and *B*, were carried out and the results are summarized in table 2.

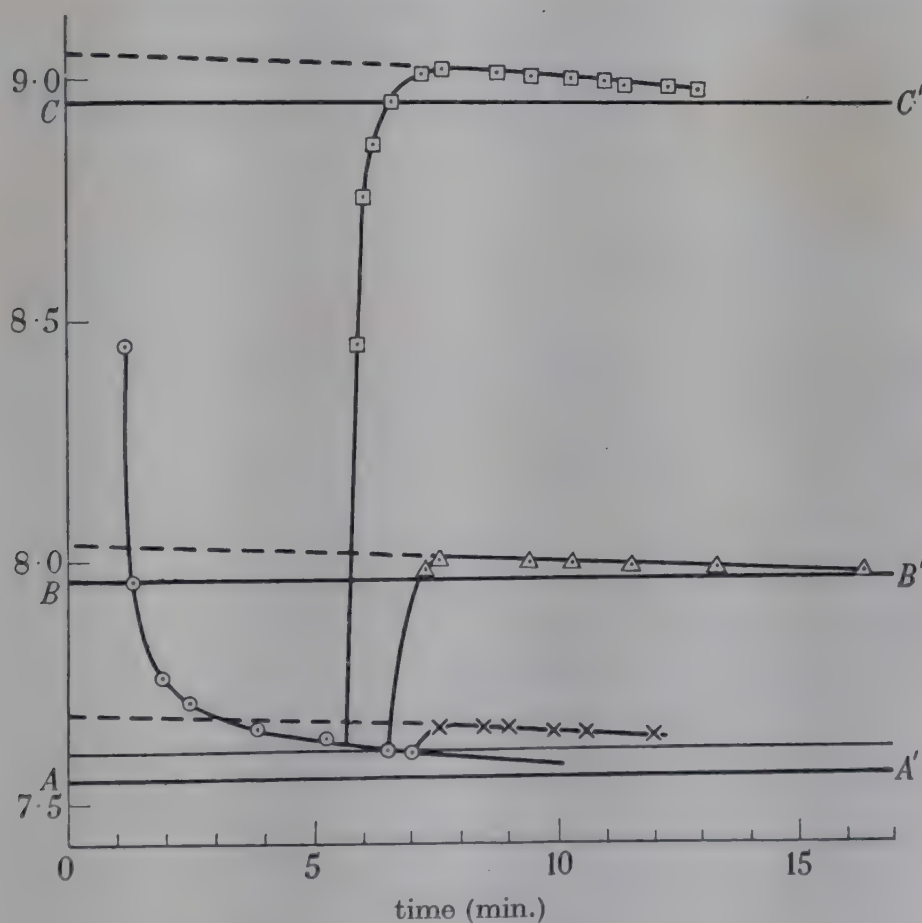


FIGURE 5. The effect of flashing on the accommodation coefficient at different wire temperatures, *AA'*, *BB'* and *CC'*.

These results show the maximum deviation of the wire resistance from normal, after flashing, to be practically constant. Consequently the change in heat loss and calculated accommodation coefficient is relatively much greater for measurements

at low- than at high-wire temperatures, contrary to Roberts's hypothesis. Evidently there is a persistence of the effects of flashing unconnected with a change in accommodation coefficient. ,

TABLE 2

	series A			series B		
steady temperature increment ($^{\circ}\text{C}$)	3.3	14.5	45.6	2.87	14.5	46.8
preselected resistance (Ω)	7.607	7.961	8.930	7.602	7.960	8.961
maximum increase after flashing (Ω)	0.038	0.033	0.051	0.047	0.043	0.064
wire temperature at zero time by extrapolation ($^{\circ}\text{C}$)	6.07	17.2	48.5	6.06	17.67	50.53
extrapolated accommodation coefficient	0.043	0.148	0.158	0.043	0.142	0.154

(c) *The effect of the time interval allowed to lapse, after flashing, on the deviation from normal*

Experiments 1 (a) and 1 (b) differ fundamentally in the time interval, after flashing, which is allowed to elapse before a measurement is made. In this experiment, the predetermined current was introduced at from 1 to 15 min. after flashing. The results

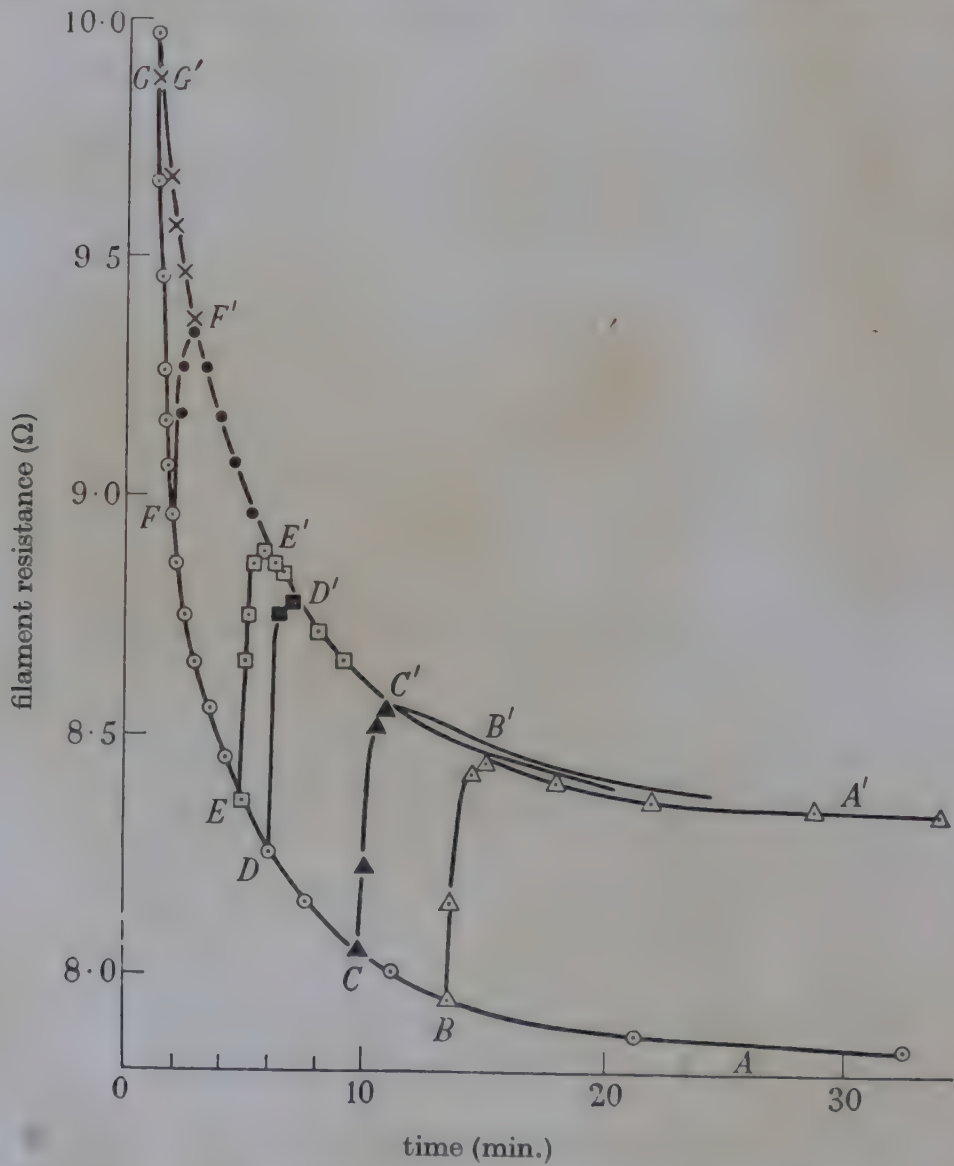


FIGURE 6. The variation in the time interval after flashing and its effect on the resistance increment.

obtained by working with the silvered vessel at a helium pressure of 0.00483 cm. are shown in figure 6. The cooling curve, *G* to *A*, is measured, as before, using a current of about 0.002 A. The predetermined current of 0.0205 A in repeat experiments is introduced at successively higher points *A* to *G* on this cooling curve to give the characteristic rise and fall in resistance. There is clearly nothing specific about the Mann (1934) procedure in which the predetermined current is passed when the pre-selected resistance is reached, after flashing. The value of the extrapolated accommodation coefficient is arbitrarily dependent on the time interval after flashing (see table 3). Mann's (1934) values are lower than those of Roberts (1930) because he worked at shorter time intervals. It is clear from figure 6 that the falling resistance

TABLE 3

accommodation coefficient	0.16	0.095	0.04	0.03	0.015
time interval (min.)	60	13.5	6	5	2

curves as *B'A'*, *C'B'* and *D'C'*—similar to those of figures 3 and 4—represent parts on a (upper) cooling curve *G'* to *A'* in which the steady resistance is the preselected one. The effects obtained in experiments as 1 (*a*) and 1 (*b*) could consequently be estimated by fixing the two (lower and upper) cooling curves *A* to *G* and *A'* to *G'*. The results of figure 6 show linear extrapolation not to be justified and throw light on the controversy between Roberts (1935) and Mann (1934) as to the best method of extrapolation (see Michels 1937).

(d) *The effect on one filament caused by flashing a similar filament mounted in the same vessel*

If the after-effects of flashing are largely thermal in nature, they should also be observed if a source of heat independent of the filament is used, as by mounting two almost identical filaments in the same vessel. The platinized vessel was used in this experiment and potential leads were connected to the external electrodes of the wire under examination. This will be termed the 'experimental' and the other the 'auxiliary' filament. Two series of experiments were carried out in the manner of 1 (*b*) at flashing temperatures of 1300 and 2900° K with a helium pressure of 0.00466 cm. The resistance-time relations after flashing the experimental and auxiliary wires to 1300 for 15 min. are shown in figure 7, curves *A* and *B* respectively. While the temperature of the experimental filament rose to only 56° C when the auxiliary wire was being flashed yet the *A* and *B* curves are similar but for a displacement in time. The resistance increment above normal, 0.113Ω, when the auxiliary wire was flashed for 15 min. was much greater than that, 0.059Ω, given by flashing the experimental filament for only 1 min. Flashing temperatures of 2900° K gave similar results, but the resistance increments were some five times greater (see table 4).

Resistance increments measured at the same time interval show the effect produced by flashing the auxiliary wire to be 80 to 85 % that given by the experimental wire. Evidently the thermal effects persist not mainly on the filament mountings but on the internal walls of the vessel. The additional contribution of the filament

leads was demonstrated by flashing the experimental wire for 1 min. at 2900° K after the auxiliary wire had been flashed at this temperature for 15 min. The resistance increment became 0.781 Ω compared with the 0.7 Ω obtained when only the auxiliary filament was flashed.

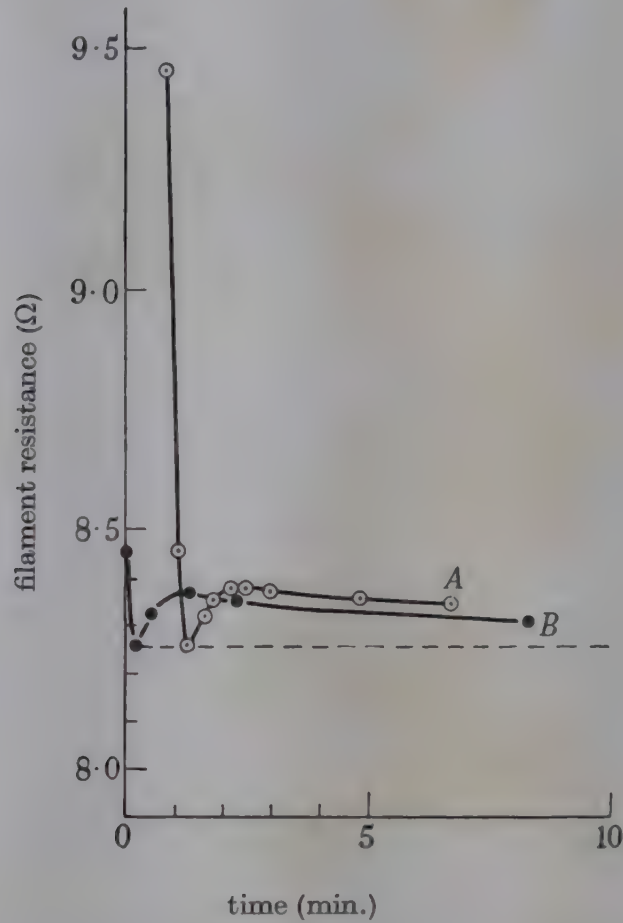


FIGURE 7. The effect of flashing the experimental wire (*A*) and the auxiliary wire (*B*) on the resistance increment.

TABLE 4

flashing conditions				maximum temp. of experimental wire (° K)	resistance maxima		
experimental wire		auxiliary wire			time increment		magnitude
(° K)	(min.)	(° K)	(min.)		(min.	sec.)	(Ω)
1300	15	—	—	1300	2	32	0.125
—	—	1300	15	329	1	22	0.113
1300	1	—	—	1300	2	23	0.059
2900	15	—	—	2900	7	20	0.686
2900	15	—	—	2900	7	39	0.713
—	—	2900	15	~ 723	7	37	0.603
—	—	2900	15	~ 723	7	31	0.574
2900	1	—	—	2900	3	55	0.501

(e) *The flashing effects in vacuo*

If, after flashing, the rate of cooling of the wire as the bath temperature is approached is governed mainly by the cooling of the wire surroundings, then effects similar to those already demonstrated should occur in the absence of gas.

The silvered vessel was used in an experiment similar to that of 2 (c). The effects produced after flashing the wire for 1 min. at 2800° K are shown in figure 8. The results are similar to those of 2 (c) but for the slower resistance change in the absence of gas.

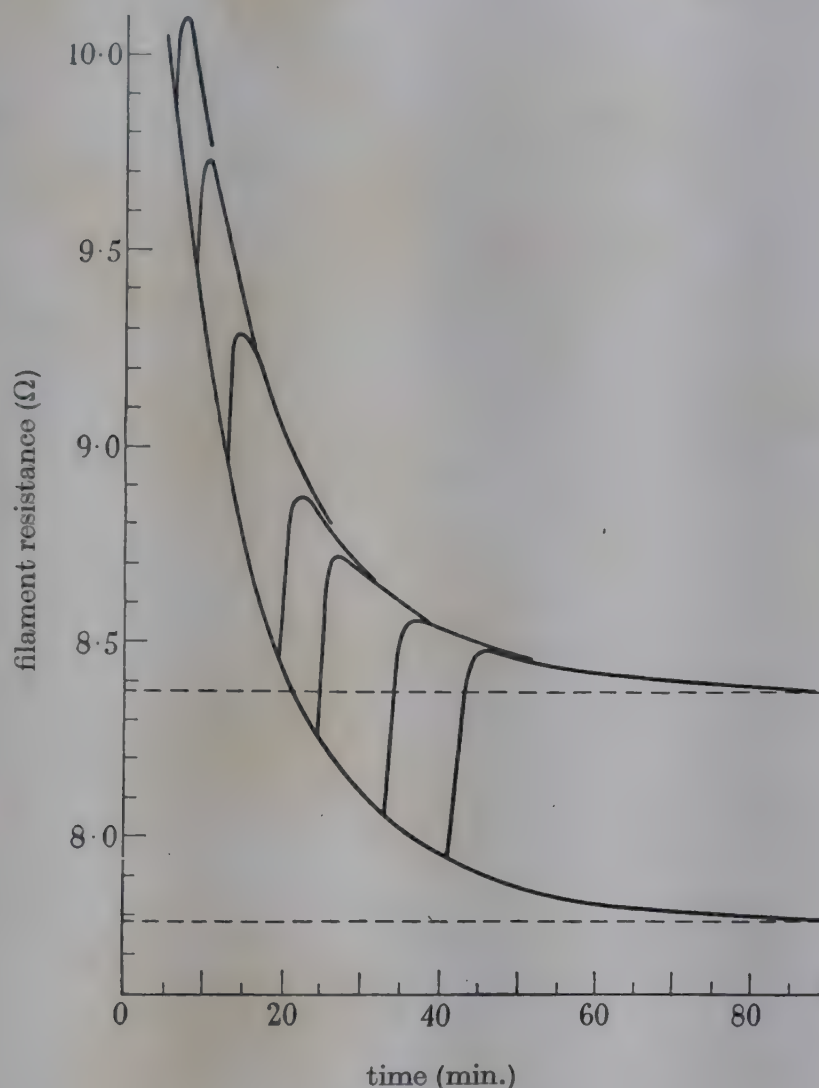


FIGURE 8. The flashing effect *in vacuo*.

The almost equivalent effects given on flashing the auxiliary and experimental wires *in vacuo*—a repetition of 2 (d)—are shown by the results in table 5.

TABLE 5

flashing conditions				maximum temp. of experimental wire (° K)	resistance maxima		
experimental wire		auxiliary wire			time increment		magnitude (Ω)
(° K)	(min.)	(° K)	(min.)		(min.)	(sec.)	
1300	15	—	—	1300	4	13	0.096
—	—	1300	15	369	3	36	0.106
1300	1	—	—	1300	4	3	0.026
2900	15	—	—	2900	15	49	0.454
—	—	2900	15	~ 773	16	10	0.455
2900	1	—	—	2900	5	24	0.175

A comparison of the results in tables 4 and 5 shows the experimental wire to be raised to a higher temperature by radiation from the auxiliary wire *in vacuo* than with helium present.

(f) *Flashing effects in vacuo with carbon and metal filament lamps*

A simple apparatus, with a Wheatstone bridge, was assembled to examine the effect of flashing carbon and metal filament lamps *in vacuo*.

The carbon lamp was a well-used B.T.H. product (100 V; 50 c.p.) which was first degassed for 50 hr. with an over-voltage of 7 %. Carbon has a negative temperature coefficient, the experimentally determined value being $\frac{1}{R_{7.8}} \frac{dR}{dT} = -0.0011 \Omega \text{deg.}^{-1}$.

The experimental procedure was that given in 1 (b). After flashing for 1 min., by connexion to the 105 V mains supply, the rise in resistance, on cooling, was followed until at 26.5° C above the bath temperature the predetermined current of 0.021 A was passed. The temperature rose above this previous steady value—indicated by a fall in resistance—and then slowly fell (see figure 9).

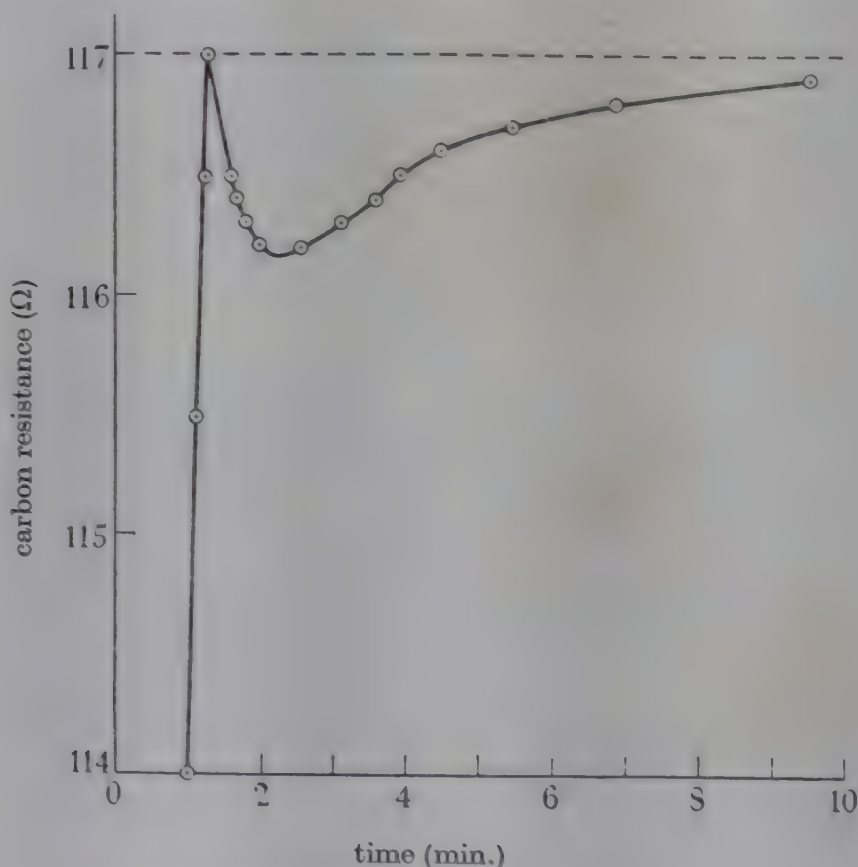


FIGURE 9. The flashing effect *in vacuo* with a carbon lamp.

The tungsten lamp was of Osram make and had a cage-wound filament (100 V; 40 W). After degassing, experiments similar to 1 (b) were carried out with flashing voltages of from 10 to 105 passed for 1 min. The fall in resistance after flashing was rapid and the resistance increments large (figure 10). Instead of a steady temperature increment of 37° C the passage of 0.0122 A, after flashing at 105 V, caused the filament to reach a transient temperature 75° C above that of the bath. Such

large effects are presumably due to the heating and subsequent slow cooling of the complex filament supports used in this type of lamp.

The results given in these experiments are of interest in that they show those previously obtained not to be dependent on the nature of the filament vessel or type of bridge circuit used.

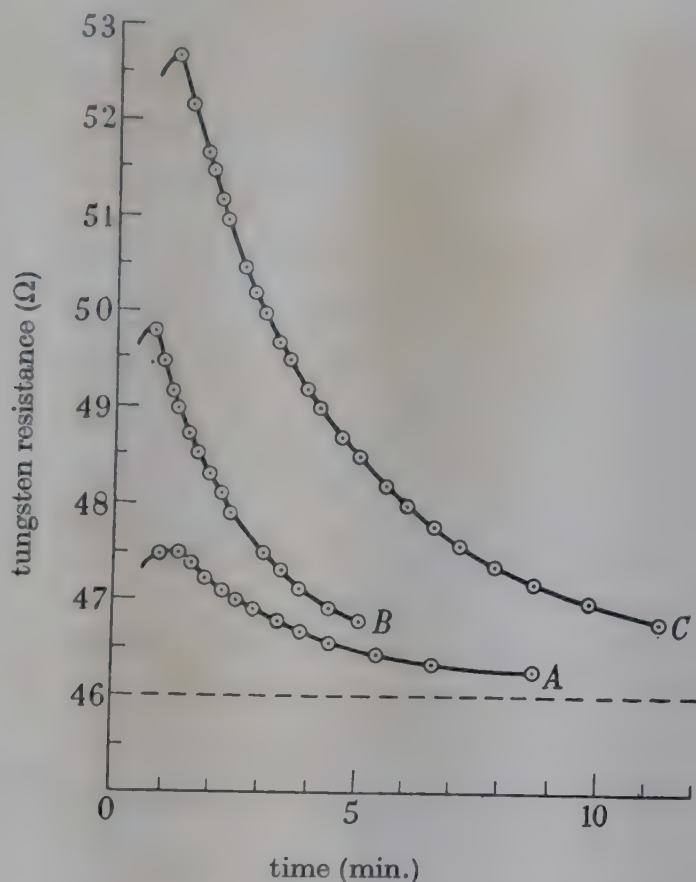


FIGURE 10. The flashing effect *in vacuo* with a tungsten lamp.
Flashing voltages: A, 10; B, 30; C, 80.

CONCLUSION

During the determination of an accommodation coefficient the temperature of the wire is raised about 20°C above that of the bath. This causes a minute increase in the temperature of the filament leads and of the internal walls of the containing vessel. During flashing, the leads and walls increase in temperature but to an extent small compared with that of the filament itself. When the flashing current is switched off the filament and its surroundings begin to cool, and it is only when the latter reaches its former steady state that heat-loss measurements can be simply correlated with the accommodation coefficient. That the relatively slow cooling of the surroundings was the limiting factor in this work was shown by:

(i) The dependence of the effect of flashing on the temperature and duration of flashing.

(ii) The substantial independence of the effect on the temperature at which the accommodation coefficient was measured.

(iii) The increase in the effect with the presence of a platinum and a silver film on the internal walls of the vessel; both conserve the heat generated by flashing.

(iv) The similarity in the effect obtained by flashing an auxiliary filament with that given by flashing the actual filament on which measurements were taken.

(v) The increased effect observed by working in the absence of a gas.

That Mann (1934) obtained even lower values than Roberts (1930) is due to the shorter cooling times he allowed to elapse before measurements were begun.

The author wishes to thank the Department of Scientific and Industrial Research for a grant, and to acknowledge the assistance of Mrs W. C. G. Bremner in the preparation of this paper. He is especially grateful to Mr D. L. Chapman, F.R.S., and (the late) Mr H. J. George for their encouragement and many helpful suggestions throughout this work.

REFERENCES

- Amdur, I., Jones, M. M. & Pearlman, H. 1944 *J. Chem. Phys.* **12**, 159.
Beeck, O. 1936 *J. Chem. Phys.* **4**, 680.
Bonhoeffer, K. F. & Rowley, H. H. 1933 *Z. phys. Chem. B*, **21**, 84.
Chapman, D. L. & Hall, W. K. 1929 *Proc. Roy. Soc. A*, **124**, 478.
Farkas, A. 1931 *Z. phys. Chem. B*, **14**, 371.
Hughes, D. R. & Bevan, R. C. 1928 *Proc. Roy. Soc. A*, **117**, 101.
Johnson, M. C. 1930 *Proc. Roy. Soc. A*, **128**, 444.
Jones, H. A. 1926 *Phys. Rev.* **28**, 202.
Knudsen, M. 1911a *Ann. Phys., Lpz.*, **34**, 593.
Knudsen, M. 1911b *Ann. Phys., Lpz.*, **36**, 871.
Knudsen, M. 1930 *Ann. Phys., Lpz.*, **6**, 129.
Langmuir, I. & Blodgett, K. B. 1932 *Phys. Rev.* **40**, 78.
Mann, W. B. 1934 *Proc. Roy. Soc. A*, **146**, 776.
Mann, W. B. & Newell, W. C. 1937 *Proc. Roy. Soc. A*, **158**, 397.
Michels, W. C. 1932 *Phys. Rev.* **40**, 472.
Michels, W. C. 1933 *Phys. Rev.* **44**, 604.
Michels, W. C. 1937 *Phys. Rev.* **52**, 1067.
Raines, B. 1939 *Phys. Rev.* **56**, 690.
Roberts, J. K. 1930 *Proc. Roy. Soc. A*, **129**, 146.
Roberts, J. K. 1932 *Proc. Roy. Soc. A*, **135**, 192.
Roberts, J. K. 1933 *Proc. Roy. Soc. A*, **142**, 518.
Roberts, J. K. 1935 *Proc. Roy. Soc. A*, **152**, 445.
Rolf, P. 1944 *Phys. Rev.* **65**, 185.
Rowley, H. H. & Evans, W. V. 1935 *J. Amer. Chem. Soc.* **57**, 2059.
Soddy, F. & Berry, A. J. 1910 *Proc. Roy. Soc. A*, **83**, 254.
Soddy, F. & Berry, A. J. 1911 *Proc. Roy. Soc. A*, **84**, 576.
Spivak, G. V. & Zacharjin, G. 1936 *Phys. Z. Sowjet.* **10**, 495.
Spivak, G. V. & Zaitzev, A. A. 1935 *C.R. Acad. Sci. U.R.S.S.* **2**, 118.
Thomas, L. B. & Olmer, F. 1943 *J. Amer. Chem. Soc.* **65**, 1036.

The thermal accommodation coefficients of gases. II. Determinations at room and at liquid-oxygen temperatures

BY J. G. M. BREMNER, *Jesus College, Oxford*

(Communicated by D. L. Chapman, F.R.S.—Received 9 July 1949)

The accommodation coefficients of helium, neon, argon, hydrogen and oxygen, on a tungsten surface, are smaller at liquid-oxygen than at room temperature. The accommodation coefficients of the inert gases varies, at both temperatures, approximately as the square root of the atomic weights. The increase in the accommodation coefficient of helium, with the hot-working of the wire observed by Roberts (1930), has not been confirmed for heating times up to 30 hr.

INTRODUCTION

In part I it was demonstrated that accommodation coefficients, measured soon after a wire had been flashed, were too small because of the persistence of the thermal effects of flashing. In consequence, the measurements of Roberts (1930) on the heat loss from a tungsten wire immersed in helium and neon have been redetermined. Because of their bearing on the mechanism of energy exchange at an interface, experiments were carried out at room and at liquid-oxygen temperatures. The work was extended to include argon and the two diatomic gases, hydrogen and oxygen. It was a further object of this work to examine whether the accommodation coefficient was dependent on the previous thermal treatment of the filament in the manner reported by Roberts (1930).

EXPERIMENTAL

The apparatus and general procedure was that described in part I. Neon and argon were obtained from the British Oxygen Co. Magnetic valves were used in their manipulation in the same manner as for helium. While the neon was purified by circulation over cooled charcoal, the method finally adopted for argon comprised its contact with a mass of copper powder and its passage over a glowing tungsten filament which had previously been thoroughly degassed *in vacuo*. The freshly reduced copper mass was at room temperature while the argon was circulated over it. It was shown in a separate experiment that oxygen was rapidly absorbed in this way. Hydrogen was admitted to its storage vessel by diffusion through an electrically heated palladium tube. Oxygen was prepared by warming crystals of potassium permanganate which had been carefully purified by three crystallizations from distilled water. Before use, the crystals were heated to ensure oxidation of the main body of organic matter present. Any condensible gases evolved after the preliminary treatment were removed by allowing the gas to stand for at least 12 hr. in a trap cooled in liquid air. The only tap in the system was the main isolation one on the pumps. The entire apparatus was well flamed before measurements were begun.

RESULTS

(1) *The effect of the prolonged heating of tungsten on the magnitude of the accommodation coefficient*

Before starting a series of measurements, it is necessary to examine whether the accommodation coefficient is dependent on the previous heat treatment of the wire. Roberts (1930) has mentioned the prolonged heating of a tungsten wire at 2000°C to increase the value of both the steady and the extrapolated accommodation coefficient (see table 1).

TABLE 1

heating time at 2000°C	accommodation coefficient (He on W)	
	steady	extrapolated
1 hr.	0.19	0.07
'many' hours	0.40	0.12
'considerable period'	0.55	0.18

The increase in the accommodation coefficient he ascribed to an increase in the roughness of the surface (cf. Loeb 1934, p. 335). Such an effect of hot-working does not appear to have been remarked by others. Indeed, Mann (1934) flashed a platinum wire for 27 hr. and subsequently observed lower values of the extrapolated accommodation coefficient than Roberts (1930). During the work described in part I the steady value of the accommodation coefficient remained at 0.16 to 0.17 in spite of the repeated flashing of the wire. An independent experiment confirmed this observation, since no change in the value of the steady accommodation coefficient was obtained after flashing for 30 hr. at 2000°C *in vacuo*. The measurements which follow need consequently not be corrected for roughness in the Roberts sense as they were carried out after a degassing period *in vacuo* of 6 hr. at 2500°K .

(2) *The accommodation coefficient of helium, neon, argon, hydrogen and oxygen at a tungsten surface in the region of room temperature*

The filament had a length of 37.3 cm. between the potential leads. Its diameter, measured by weighing specimens of the hot-worked wire and also by using the specific resistance data of Jones (1926), varied from 0.0059 to 0.0062 cm., which is close to the figure 0.006 cm. given by the suppliers (Johnson & Matthey). The value 0.0061 cm. was used in subsequent calculations. Heat-loss measurements were taken at a temperature increment 22.8°C above that of the bath which was at 30°C . The heat loss *in vacuo* at this temperature increment was $0.00032\text{ cal. sec.}^{-1}$, one-fifth to one-tenth of that with the gas present. In deriving an accommodation coefficient, Langmuir & Blodgett (1932) use as the incident gas temperature that calculated at one free path distance from the wire. As the wire-wall distance in this work varied from less than one path length at the top of the filament to about twice the free path at the bottom, the gas temperature was taken to be that of the wall (cf. Kennard 1938, p. 320). Heat-loss measurements on the gases were taken 1 hr. after the filament had been flashed for 10 min. *in vacuo* at 2500°K (table 2). In deriving accommodation coefficients for hydrogen and oxygen, the equation for the theoretical heat loss given in part I was modified in the normal Knudsen (1911a) manner

by introducing the expression $(\frac{1}{2} + C_v/2R)$, in which C_v is the specific heat of the gas at constant volume and R the gas constant per mole. This expression implies the accommodation coefficient for the transfer of internal energy to be the same as that for translational energy, and that a Maxwellian distribution is valid for both forms of energy (Loeb 1934, p. 322). Knudsen's (1930) experimental confirmation of this equality has been criticized by others (Bonhoeffer & Rowley 1933; van Wijk 1932; Shafer, Rating & Eucken 1942).

TABLE 2. THE ACCOMMODATION COEFFICIENT AT A WIRE TEMPERATURE OF 52.8°C

gas	pressure (cm. Hg)	heat loss in	accommodation coefficient
		gas/cm. pressure (cal. sec. ⁻¹)	
helium	0.00670	0.17	0.16
neon	0.00657	0.16	0.33
argon	0.00372	0.22	0.64
hydrogen	0.00280	0.58	0.26
oxygen	0.00500	0.42	0.72

(3) *The accommodation coefficient of helium, neon, argon, hydrogen and oxygen at liquid oxygen temperature*

The various factors involved in the determination of an accommodation coefficient at low temperatures are treated separately.

(a) *Pressure measurement*

Because of thermal transpiration the pressure measured on the McLeod gauge will be higher than the gas pressure in the cold-filament vessel. The magnitude of the effect depends on the ratio of the tube diameter at the temperature interface, d , to the mean free path of the gas molecules, λ . Theoretical treatments have been given by Reynolds (1879) for $\lambda \gg d$, by Knudsen (1910) for $\lambda < d$ (cf. Roberts 1932) and by Loeb (1927, p. 529; 1934, p. 362) and Knudsen (1927) when $\lambda \approx d$. In this work d/λ varied between 1 and 4. The accuracy of the Loeb-Knudsen equation in this range of d/λ appears largely dependent on the relation used for finding the viscosity change with temperature. As the general equation of Sutherland breaks down at low temperatures (Jeans 1921, p. 286), it was decided to measure the extent of thermal transpiration for each of the gases under examination. The cold-filament vessel was allowed to warm to room temperature after a heat-loss measurement had been made, and from the change in the McLeod reading the pressure in the cold vessel can be derived (table 3).

TABLE 3

gas	McLeod reading		pressure in filament vessel in liquid oxygen	magnitude of thermal transpiration (cm. Hg)
	filament vessel in liquid oxygen	filament vessel at room temperature		
helium	0.00235	0.00305	0.00151	0.00084
neon	0.00286	0.00388	0.002037	0.00082
argon	0.00359	0.00520	0.002945	0.00064
hydrogen	0.00277	0.00374	0.001956	0.00081
oxygen	0.00292	0.00425	0.002415	0.00050

In accordance with theory, the effect of thermal transpiration is greater for argon and oxygen than for helium. Moreover, the magnitude of the pressure change is, as required, smaller than that given by the Reynolds equation.

(b) *Temperature measurement*

It was not possible to complete an accurate determination of the temperature coefficient of resistance at liquid-oxygen temperatures. However, as Holborn (1919) and Roberts (1932) have already shown this coefficient to be independent of temperature, in the range considered, the previous value determined at room temperature, $0.0041 \Omega \text{ deg.}^{-1}$, was used in the present work. Such a coefficient, moreover, corresponds well with the observed resistance fall of from 7.08 to 1.15Ω between room and liquid-oxygen (90° K) temperatures.

(c) *The heat loss in vacuo*

In the absence of gas, experiment showed a minute energy input to cause a considerable rise in the temperature of the wire. A temperature increment of 100° was obtained when 0.0025 A was passed. This indicates a heat loss less than 1% that observed with a gas present, and consequently no correction was applied. Roberts (1932) came to the same conclusion.

(d) *The theoretical heat loss*

While the specific heat of the inert gases remain unchanged at $3R/2$ (*Int. Crit. Tables* 5, 80), there is a slight fall in the specific heat of oxygen to 4.91 at liquid-oxygen temperatures (Scheel & Heuse 1913; *I.C.T.* 5, 80), where the value for hydrogen falls to 3.26 cal./mole (Scheel & Heuse 1913; *I.C.T.* 5, 82). Although the specific heat of hydrogen is changing rapidly at liquid-oxygen temperatures it is unnecessary to correct for the difference in the heat capacity of the incident and reflected gas molecules because of the small value of the accommodation coefficient.

(e) *The determination of accommodation coefficients*

Before making a determination, the wire was flashed *in vacuo* at 2500° K for 3 min. with the vessel surrounded by liquid oxygen. Gas was admitted 20 to 30 min. later and measurements taken 1 hr. after the flashing current had been switched off. Current and resistance readings were taken at successively higher temperatures of the wire. The heat loss-resistance relation was not quite linear (cf. Langmuir & Blodgett 1932) (see figures 1 and 2). Because of this non-linearity, the accommoda-

TABLE 4. THE ACCOMMODATION COEFFICIENT AT A WIRE
TEMPERATURE 25° ABOVE THAT OF THE BATH

gas	pressure (cm. Hg)	heat loss in gas/cm. pressure (cal. sec. $^{-1}$)	accommodation coefficient
helium	0.00151	0.09	0.041
neon	0.002037	0.08	0.081
argon	0.002945	0.11	0.16
hydrogen	0.00190*	0.29	0.09
oxygen	0.002415	0.23	0.20

* At the lower bath temperature of 87.5 (cf. table 3).

tion coefficients were derived in two ways. The first was the normal one, in which the heat loss at a given temperature increment— 25° in these calculations—was compared with the theoretical. The results obtained in this way are given in table 4.

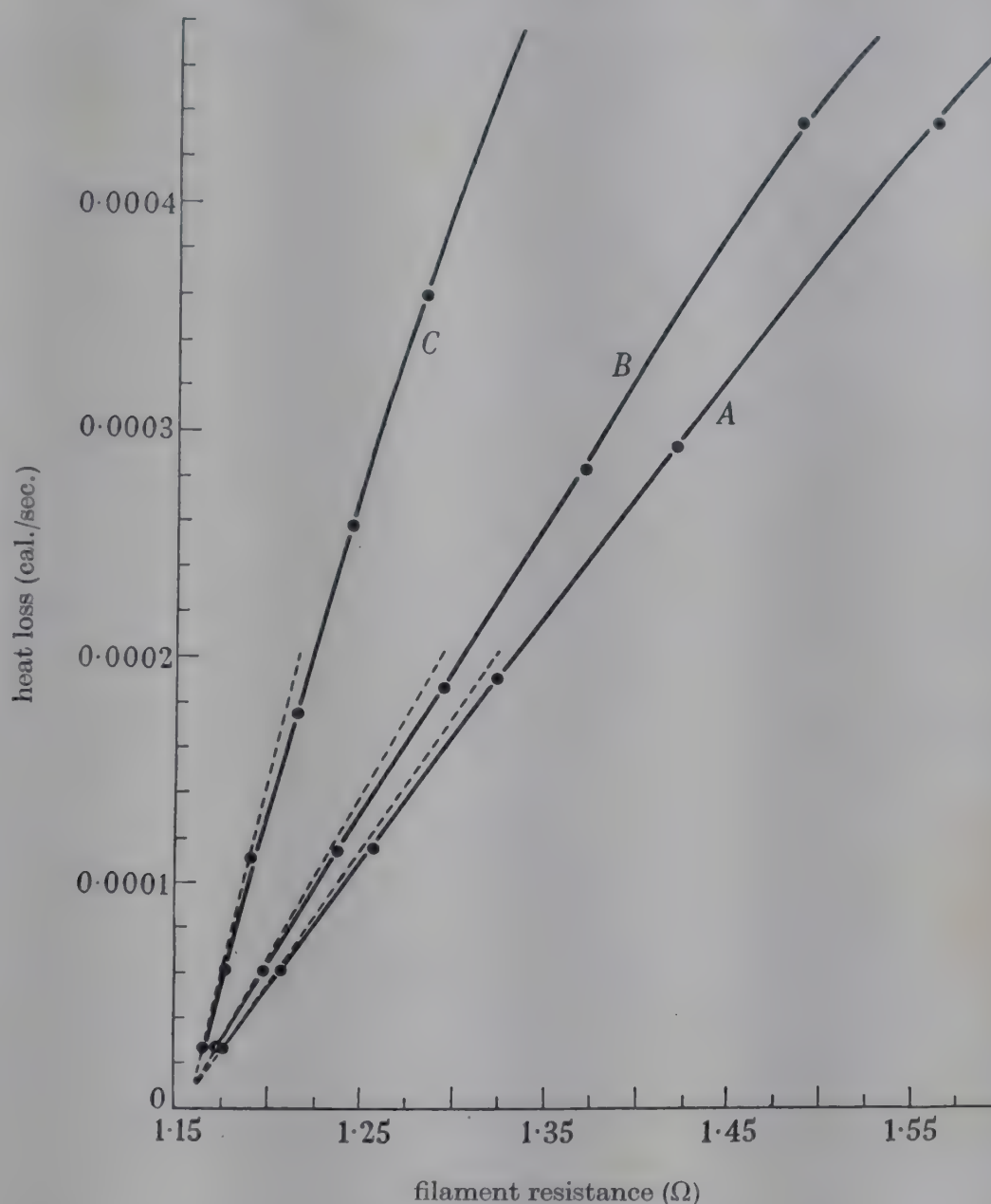


FIGURE 1. The heat loss at increasing temperatures. A, helium; B, neon; C, argon.

In the second method, the tangents to the heat-loss resistance curves were drawn (see figures 1 and 2) and their slopes, $d(\Delta H_E)/d(\Delta T)$, compared with the theoretical ones, $d(\Delta H_c)/d(\Delta T)$, where ΔT is the temperature increment between the incident gas and the wire. For a given gas,

$$\Delta H_c = Z\Delta T, \quad \text{where } Z \text{ is a constant,}$$

while

$$\begin{aligned} \Delta H_E &= b(R_T - R_0), \quad \text{where } R_0 \text{ and } b \text{ are constants and } R_T \text{ is the wire resistance,} \\ &= 0.0041bR_0\Delta T. \end{aligned}$$

Consequently, the accommodation coefficient, a , is given by

$$a_{\Delta T \rightarrow 0} = \frac{d(\Delta H_E)}{d(\Delta T)} \bigg/ \frac{d(\Delta H_c)}{d(\Delta T)} = \frac{0.0041bR_0}{Z}.$$

Owing to the slight curvature of the heat loss-resistance relation (cf. Langmuir & Blodgett 1932) accommodation coefficients calculated in this way will be slightly

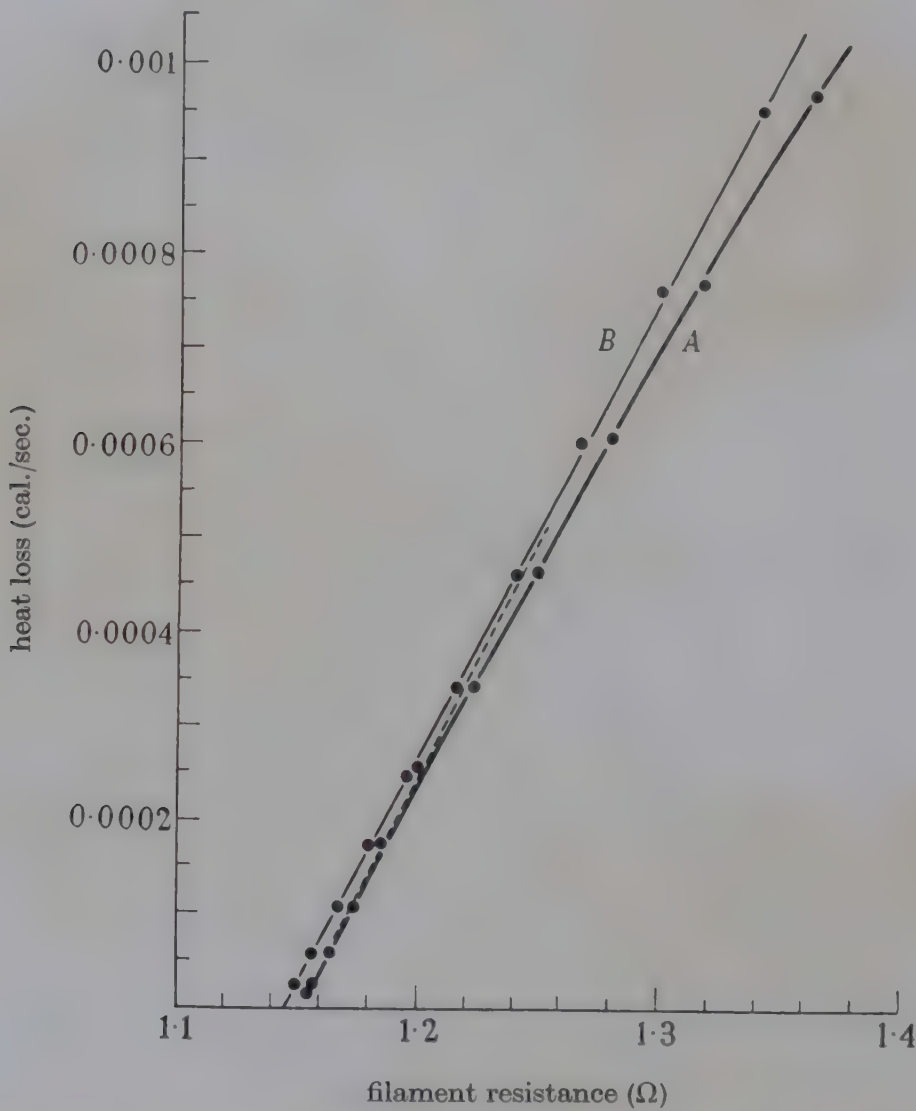


FIGURE 2. The heat loss at increasing temperatures. *A*, oxygen; *B*, hydrogen.

TABLE 5. THE ACCOMMODATION COEFFICIENT OF HELIUM, NEON, ARGON, HYDROGEN AND OXYGEN AT LIQUID-OXYGEN TEMPERATURE

gas	slope of heat loss against temperature increment		accommodation coefficient
	observed	theoretical	
helium	0.57×10^{-5}	13.21×10^{-5}	0.043
neon	0.69×10^{-5}	7.95×10^{-5}	0.087
argon	1.69×10^{-5}	8.17×10^{-5}	0.21
hydrogen	2.23×10^{-5}	25.3×10^{-5}	0.09
oxygen	2.35×10^{-5}	11.1×10^{-5}	0.21

larger than those given by the normal method (cf. Thomas & Olmer 1943). Among the inert gases, the difference will be greatest with argon—which has the highest heat loss—as the results in table 5 show.

These results show the accommodation coefficients at liquid-oxygen temperature to be $\frac{1}{3}$ to $\frac{1}{4}$ their values at room temperature.

DISCUSSION

Since they were derived in similar fashion, the results in tables 2 and 4 will be used for comparative purposes. The comparison shows the heat conductivity of both the mono- and diatomic gases at liquid-oxygen temperatures to be half that at room temperature, i.e. proportional to the square root of the temperature. As the theoretical heat loss is inversely proportional to the square root of the temperature, the accommodation coefficient becomes approximately proportional to the temperature. Taking into consideration the specific-heat term (cf. Beeck 1936, 1937) $(\frac{1}{2} + C_v/2R) = 1 + C_v$, the dependence of the accommodation coefficient, α , on temperature is given by $\alpha \propto \frac{T^{1.3}}{(1 + C_v)}$, where T is the wire temperature.

Values calculated in this way are compared with those determined experimentally in table 6.

TABLE 6. THE ACCOMMODATION COEFFICIENT AT LIQUID-OXYGEN TEMPERATURES

	He	Ne	Ar	H ₂	O ₂
α observed	0.041	0.081	0.16	0.09	0.20
α calculated	0.041	0.085	0.165	0.09	0.19

But for the values of Roberts (1932) and Spivak & Zaitzev (1935) and Spivak & Zacharjin (1936), obtained after flashing, there appears to be no published data on the variation of the accommodation coefficient of the inert gases with temperature at a tungsten surface. On platinum, Grilly, Taylor & Johnston (1946) found, as in this work, that the accommodation coefficient decreases with decreasing wire temperatures. Hydrogen behaved exceptionally as its accommodation coefficient increased at lower temperatures. This observation is well substantiated (Knudsen 1911*a, b*; Bonhoeffer & Rowley 1933; Archer 1935; Gregory & Dock 1938). Langmuir & Blodgett (1932) have examined the behaviour of hydrogen at a tungsten surface. They found the accommodation coefficient to increase rapidly below room temperature only if oxygen was present. In its absence the increase was small (0.14 to 0.19); as the measurements were, however, carried out with a constant wall temperature, it is uncertain how far they can be compared with the present results. Moreover, no correction for thermal transpiration appears to have been made. The effect of temperature on the accommodation coefficient has been treated theoretically from different viewpoints by Zener (1932), Landau (1935) and Devonshire (1937). All can readily account for a fall in the accommodation coefficient with a decrease in temperature. The Landau relation gives $\alpha \propto T^{1.5}$, similar to that deduced in the present work (cf. Spivak & Zacharjin 1936).

In agreement with published data (Soddy & Berry 1910, 1911; Dickens 1933; Thomas & Olmer 1943; Amdur, Jones & Pearlman 1944), accommodation coefficients for the inert gases were found to increase with increasing atomic weight. This holds both at room and at liquid-oxygen temperatures. Approximately, the accommodation coefficient varies as the square root of the atomic weight, m (table 7).

TABLE 7. THE DEPENDENCE OF THE ACCOMMODATION COEFFICIENT ON THE MOLECULAR WEIGHT

gas	He	Ne	A
$\sqrt{(m)}/a$ at room temperature	12.5	13.5	10.1
$\sqrt{(m)}/a$ at liquid O ₂ temperature	49	55	40

Consequently, for the monatomic gases, $a \propto T^{1/3} \sqrt{m}$, as the specific heat term is independent of temperature.

The author wishes to thank the Department of Scientific and Industrial Research for a grant, and to acknowledge the assistance of Mrs W. C. G. Bremner with many of the calculations and of Miss A. A. Bremner in the preparation of this paper. He is especially grateful to Mr D. L. Chapman, F.R.S., and (the late) Mr H. J. George for their encouragement and many helpful suggestions throughout this work.

REFERENCES

- Amdur, I., Jones, M. M. & Pearlman, H. 1944 *J. Chem. Phys.* **12**, 159.
 Archer, C. T. 1935 *Phil. Mag.* **19**, 901.
 Beeck, O. 1936 *J. Chem. Phys.* **4**, 680.
 Beeck, O. 1937 *J. Chem. Phys.* **5**, 268.
 Bonhoeffer, K. F. & Rowley, H. H. 1933 *Z. Phys. Chem. B*, **21**, 84.
 Devonshire, A. F. 1937 *Proc. Roy. Soc. A*, **158**, 269.
 Dickens, B. G. 1933 *Proc. Roy. Soc. A*, **143**, 517.
 Gregory, H. S. & Dock, E. H. 1938 *Phil. Mag.* **25**, 129.
 Grilly, E. R., Taylor, W. J. & Johnston, H. L. 1946 *J. Chem. Phys.* **14**, 435.
 Holborn, L. 1919 *Ann. Phys., Lpz.*, **59**, 164.
 Jeans, J. H. 1921 *Dynamical theory of gases*. Cambridge University Press.
 Jones, H. A. 1926 *Phys. Rev.* **28**, 202.
 Kennard, E. H. 1938 *The kinetic theory of gases*. New York: McGraw Hill.
 Knudsen, M. 1910 *Ann. Phys., Lpz.*, **31**, 205.
 Knudsen, M. 1911 *Ann. Phys., Lpz.*, **34**, 593.
 Knudsen, M. 1911 *Ann. Phys., Lpz.*, **36**, 871.
 Knudsen, M. 1927 *Ann. Phys., Lpz.*, **83**, 797.
 Knudsen, M. 1930 *Ann. Phys., Lpz.*, **398**, 129.
 Landau, L. 1935 *Phys. Z. Sowjet.* **8**, 489.
 Langmuir, I. & Blodgett, K. B. 1932 *Phys. Rev.* **40**, 78.
 Loeb, L. B. 1927 *The kinetic theory of gases*. 1st ed. New York: McGraw Hill.
 Loeb, L. B. 1934 *The kinetic theory of gases*. 2nd ed. New York: McGraw Hill.
 Mann, W. B. 1934 *Proc. Roy. Soc. A*, **146**, 776.
 Reynolds, O. 1879 *Phil. Trans.* **170**, 727.
 Roberts, J. K. 1930 *Proc. Roy. Soc. A*, **129**, 146.
 Roberts, J. K. 1932 *Proc. Roy. Soc. A*, **135**, 192.
 Roberts, J. K. 1933 *Proc. Roy. Soc. A*, **142**, 518.
 Roberts, J. K. 1935 *Proc. Roy. Soc. A*, **152**, 445.

- Scheel, K. & Heuse, W. 1913 *Ann. Phys., Lpz.*, **40**, 473.
 Shafer, K., Rating, W. & Eucken, E. 1942 *Ann. Phys., Lpz.*, **42**, 176.
 Soddy, F. & Berry, A. J. 1910 *Proc. Roy. Soc. A*, **83**, 254.
 Soddy, F. & Berry, A. J. 1911 *Proc. Roy. Soc. A*, **84**, 576.
 Spivak, G. V. & Zacharjin, G. 1936 *Phys. Z. Sowjet.* **10**, 495.
 Spivak, G. V. & Zaitzev, A. A. 1935 *C.R. Acad. Sci. U.R.S.S.* **2**, 118.
 Thomas, L. B. & Olmer, F. 1943 *J. Amer. Chem. Soc.* **65**, 1036.
 van Wijk, W. R. 1932 *Z. Phys.* **75**, 584.
 Zener, C. 1932 *Phys. Rev.* **40**, 1016.

The collection of positive ions by a probe in an electrical discharge

By R. L. F. BOYD, *Department of Mathematics, University College, London*

(Communicated by H. S. W. Massey, F.R.S.—Received 8 August 1949—

Revised 2 December 1949)

The treatment of the positive ions in an electropositive gas discharge as though they had a Maxwellian velocity distribution giving a current to a negative probe equal to the random-ion current, is not consistent with most probe measurements, which on this theory would lead to the untenable conclusion that the positive ions had a temperature about equal to that of the electrons. Theoretical reasons are given for the view that the energy distribution of the positive ions at a sheath edge must comply with a certain criterion, and unless this criterion is fulfilled the probe will not be screened by a sheath but the field of the probe will penetrate into the plasma. The penetration of this field modifies the energy distribution and concentration of the positive ions in the probe neighbourhood until, if the probe voltage is sufficient, the criterion is satisfied and a sheath forms. The probe potential necessary to bring this about has been calculated for idealized high- and low-pressure cases. (It varies between 1 and 4 V). The effect of the penetrating field on the saturation positive-ion current as predicted by the conventional theory has been calculated. At pressures of the order of 1 mm. Hg and above, the saturation current to a negative probe is approximately equal to the random current of positive ions in the gas. At pressures below 0.1 mm. Hg it is considerably greater, being very nearly equal to the value the conventional theory would give if the positive ions had a temperature equal to that of the electrons.

Experiments with a small probe in the form of a grid with a collector behind it made it possible to obtain Langmuir probe characteristics in argon with the electron, ion and secondary emission components separated from one another. Although the smallness of the probe (necessary in order not to disturb the discharge) prevents the rigorous application of the theory because of edge effects, yet the general predictions of the theory are confirmed. In particular, the positive-ion current saturation is seen not to occur until the probe is several volts negative to the space, and the value of this current is low at high pressures and high at low pressures.

The retardation part of the positive-ion curve enables the determination of a 'temperature' corresponding to the transverse component of the ion velocities. At 0.004 mm. Hg it is about 0.3 of the electron temperature, and about a tenth the latter at 0.0009 mm. Hg.

Observations have also been made upon the anisotropy of the ion and electron-energy distributions, and the probe has been shown to be especially useful in studying departures of the electron-velocity distribution from truly Maxwellian form at high energies.

Secondary emission from the platinum collector by photons and metastable particles amounted to about 10 % of the saturation positive-ion current.

1. INTRODUCTION

Since its inception in 1924, the Langmuir-Mott-Smith probe has been the standard technique for the measurement of electron concentration and energy and of the space potentials in discharge plasmas. The technique has received a number of extensions (Emeléus, Greaves & Montgomery 1936; Greaves & Johnston 1936; Druyvesteyn 1930), but comparatively little study has been made of the collection of positive ions, and it is still usual to treat them as though they had an isotropic energy distribution not far from Maxwellian, although the objections to such a conception have been clearly stated by Langmuir as long ago as 1929 (Langmuir & Tonks 1929).

From the point of view of probe theory the evidence against the simple positive ion 'temperature' conception is twofold.

(i) Experimentally the ratio of the positive and negative saturation currents obtained under such a theory (neglecting secondary emission effects, which are of the order of 10 %) should be given by

$$i_e/i_+ = (T_e/T_+)^{1/2} (m_+/m_e)^{1/2},$$

where T_e , m_e and i_e are the electron temperature, mass and random current density respectively, and T_+ , m_+ and i_+ those of the positive ions.* However, to account for the ratio of saturation currents often observed, it would be necessary to assume a positive-ion temperature as great and sometimes greater than the electron temperature. An estimate of the mean positive-ion energy from the fields normally existing in the discharge would indicate that it probably varies from about $\frac{1}{100}$ of the electron temperature at the higher pressures to a rather larger fraction at those pressures for which the mean free path is of the order of the tube radius (Frisch & Kagan 1943). At still lower pressures the ion energies will decrease again as the regime becomes that of the hot-cathode low-pressure arc.

(ii) It is not possible to satisfy the theoretical conditions at the edge of a positive-ion sheath unless the ion-energy distribution fulfils certain conditions. The essence of this objection is given in a paper by Langmuir (1929) during a discussion of the double sheath at the cathode of a discharge. The very important bearing that this work of Langmuir has upon the conventional interpretation of the action of probes which are negative to the space, does not seem to have been sufficiently generally realized.

These experiments were undertaken to determine the positive-ion current to a plane probe as a function of the applied voltage (by separating it instrumentally from the electron component) in order to enable the formulation of a consistent theory of probe action at potentials negative to the space. Only electro-positive gases will be considered, but work is in hand to extend the study to gases containing negative ions.

2. THE EXPERIMENTAL METHOD

The separation of the positive ions and electrons arriving at the probe has been achieved by the use of a plane collector electrode, screened by a fine wire grid. By making the collector electrode first positive and then negative to the space potential,

* On certain theoretical grounds Langmuir suggested the form $i_e/i_+ = \frac{1}{2}(T_e/T_+)^{1/2} (m_+/m_e)^{1/2}$ to allow for the inevitable one-sidedness of the velocity distribution at the probe surface.

the electron and ion currents reaching the plane of the grid at any particular setting of the grid potential could be measured consecutively. Attempts to study the collection energies and currents of positive ions, by a similar arrangement, have been made before, notably by Langmuir and his collaborators (Tonks, Mott-Smith & Langmuir 1926) and by Sena (1935), but in these cases the large size of the screening electrode prevented it from being used at a voltage positive to the space without becoming the anode of the discharge. Accordingly, these workers put a large negative voltage on the screen and attempted to measure the positive-ion energy distribution by varying the retarding voltage on the collector. This they succeeded in doing, but as will be seen later the energies so measured arise principally from the penetration of the plasma by the field of the screen and are therefore instrumental rather than characteristic of the discharge.

The use of modern methods for building very small grids has now made it possible to operate the screen at voltages positive to the discharge and so, by varying the screen potential over a range of voltages, to obtain a true Langmuir characteristic divided into its two components. Since the positive ions reach the grid through a retarding field there is no disturbance of the measured-energy distribution by a field from the grid.

Characteristics have been taken in an argon discharge for a range of pressures from 0.75 to 0.00085 mm. Hg and for a number of probe positions and orientations, each characteristic having its electron and positive-ion components recorded separately. The frame of the grid acted at the same time as a guard ring to reduce the fringing effect due to finite sheath thickness.

3. THEORETICAL CONSIDERATIONS

In this section the theoretical aspect of the collection of positive ions is discussed. The probe is assumed to be small enough not to disturb the discharge conditions appreciably, but sufficiently large in relation to the disturbed region to be regarded as a plane and to make the problem unidimensional.

Let the probe be at a potential negative to the space. The potential V at a distance x from the probe is given by

$$d^2V/dx^2 = -4\pi e(n_+ - n_e), \quad (1)$$

n_+ and n_e being the local positive-ion and electron concentrations.

At the sheath edge space charge neutralization is almost complete, so that $n_e = \phi n_+$, where ϕ is a constant nearly equal to unity, which shall for the present define the sheath edge. At a distance δx just inside the sheath edge let the potential be δV less than at the edge:

$$\frac{d^2V}{dx^2} = -4\pi en_+^s \left\{ \frac{\int_0^\infty f(E) E^{\frac{1}{2}} dE}{\int_0^\infty f(E) (E + \delta V)^{\frac{1}{2}} dE} - \phi e^{-\delta V/V_e} \right\}, \quad (2)$$

where $f(E)$ is the distribution of the energies associated with the x -component of the positive-ion velocities at the sheath edge and $V_e = kT$, T being the electron

temperature. The superscripts 0, S and x refer to undisturbed plasma, sheath edge and position x . The appearance of $f(E)$ in both numerator and denominator is equivalent to the assumption that collisions can be neglected in the distance δx .

For a sheath to form V must increase rapidly, so

$$\left(\int_0^\infty f(E) E^{\frac{1}{2}} dE \right) / \left(\int_0^\infty f(E) (E + \delta V)^{\frac{1}{2}} dE \right) > e^{-\delta V/V_e}, \quad (3)$$

which if δV is very small gives the 'sheath criterion'

$$\left(\int_0^\infty f(E) E^{\frac{1}{2}} dE \right) / \left(\int_0^\infty f(E) E^{-\frac{1}{2}} dE \right) > \frac{1}{2} V_e. \quad (4)$$

This sets certain restrictions upon the form of the positive-ion energy distribution at the sheath edge. It cannot be satisfied by the one-sided Maxwellian distribution normally assumed. Unless the criterion is satisfied the field from the probe will penetrate the plasma, this penetration eventually modifying the positive-ion energy distribution to comply with the criterion.

Consider this extra-sheath region into which the sheath field penetrates. Since $\phi \simeq 1$, to a high degree of approximation the positive-ion density is given by

$$n_+^x = n_+^0 e^{-V/V_e}. \quad (5)$$

(The potential where the superscript 0 applies is taken as zero.) The potential distribution must so adjust itself to comply with (5) and to give an energy distribution at the sheath edge fulfilling (4).*

It is impossible to satisfy (5) if the ions fall freely in the extra sheath field unless either ionization in the extra sheath region is taken into account or the probe is sufficiently small for the field to be distorted by the geometry. We distinguish then three regimes under which (5) may be satisfied.

- (i) The high-pressure regime (order 1 mm. Hg) where the ion flow is diffusive.
- (ii) The low-pressure regime, where the extra-sheath field penetration continues until ion production in the region by the normal discharge processes becomes sufficient to permit the satisfying of (5) and (4).
- (iii) The case of a very small probe or cylindrical or spherical probes. These will not be considered here. Langmuir (1929) has treated the cylindrical probe at low pressure.

The high-pressure case

If, as is the case for the inert gases, the cross-section for charge exchange for the positive ions is larger than the gas kinetic cross-section each ion may be assumed to start from rest from the position of its production in a charge exchange process. This is the mobility mechanism described by Sena (1946) and it leads to a particularly simple treatment of the mobility.

* There is an important difference between Langmuir's attitude to a plasma boundary and that given here. Langmuir (1929) abandons his plasma equation, on approaching the boundary, when the approximations made in solving it cease to be valid. At this point he starts his sheath equation and uses a mean value for the ion-energy distribution. It seems better, however, to make use of the sheath criterion in defining the sheath edge, as the potential and extent of the boundary region depends upon the form of $f(E)$.

I will digress here to define the region of disturbance by the penetrating field (the extra-sheath region) as the region between the surface at which the sheath criterion is satisfied and the surface at which the mean normal component of the velocities of the positive ions is equal to the velocity they would have acquired in the penetrating field if they were assumed to have started from rest. This surface may be approximated to in the case of a probe parallel to the plasma field at high pressures by the surface at which the penetrating field becomes equal to the plasma field which existed in the absence of the probe.

If the mean free path for charge exchange be λ and there are a number of ions drifting in a field X , the probability of a particle having an energy E is

$$\frac{1}{\lambda X} e^{-E/\lambda X} dE = f(E) dE \quad (6)$$

The criterion of sheath formation then becomes

$$\lambda X > V_e. \quad (7)$$

(The approximation has been made that the field is constant over a mean free path. The effect of the variation of the field over a mean free path, and of the variation of charge exchange cross section with energy will be to introduce a numerical factor of order unity into (7)).

Let v_+^0 be the mean normal velocity with which ions enter the disturbed region. Let the potential at the boundary of entry be zero and at a point distant x inside V_x , where also the drift velocity is v_+^x .

Combining (5) with Sena's mobility equation $v_+^x = \mu \left(\frac{dV}{dx} / p \right)^{\frac{1}{2}}$, where μ is Sena's 'mobility' constant and p is the pressure, and noting that $n_+^x v_+^x = n_+^0 v_+^0$, we obtain on integrating

$$V_x = -\frac{1}{2} \log \left\{ 1 - \frac{2}{V_e} \left(\frac{v_+^0}{\mu} \right)^2 p x \right\}. \quad (8)$$

This distribution of potential between the sheath edge and the edge of the disturbed region, where it merges with the main field of the discharge, is shown in figure 1a.

Introducing (7) into the differential form of (8) we obtain an expression for the potential of the sheath edge

$$\lambda p (v_+^0 / \mu)^2 e^{2\Delta V / V_e} = V_e, \quad (9)$$

yielding if $\lambda p = \beta$ (say)

$$\Delta V = \frac{1}{2} V_e \log \{ (V_e / \beta) (\mu / v_+^0)^2 \}. \quad (10)$$

ΔV is an important quantity. It is the difference in voltage between space potential (at which electron current saturation occurs) and the potential at which a positive-ion sheath forms resulting in positive-ion current saturation. Over this range of voltage the probe is not covered by a sheath. (The fact that in practice saturation does not occur until the sheath forms shows that the current entering the disturbed region is not independent of its extent.)

Equations (8) and (10) combine to give the extent d of the disturbed region

$$d/\lambda = (V_e/2\beta) (\mu/v_+^0) - \frac{1}{2}. \quad (11)$$

In this regime λ is small and therefore d is small, so that ionization in the disturbed region will not contribute to the probe current and the positive-ion saturation current will equal that entering the disturbed region $-en_+^0v_+^0$.

It is reasonable to assume $v_+^0 = \mu(X_L/p)^{\frac{1}{2}}$, where X_L is the axial field in the tube. The ratio of the saturation currents is then

$$\left(\frac{i_e}{i_+}\right)_{\text{sat.}} = \frac{n_e^0(e/m_e \cdot V_e/2\pi)^{\frac{1}{2}}}{n_+^0\mu(X_L/p)^{\frac{1}{2}}} = 1.62 \times 10^7 \left(\frac{V_e}{X_L/p}\right)^{\frac{1}{2}}. \quad (12)$$

(Here, as elsewhere, where numerical values are introduced, practical units are used and p is in mm. Hg.)

Figure 1a shows that the greatest contribution to ΔV occurs in a short distance near the probe, whereas the value of $(i_e/i_+)_{\text{sat.}}$ is very sensitive to ionization in the disturbed region and to v_+^0 . Equation (10) should therefore be valid to a lower pressure than that at which (12) fails.

For $X_L/p < 0.75$ V/mm. Hg ion energies are roughly thermal, and it is better to take $v_+^0 = (2\pi)^{-\frac{1}{2}}kT$, where T is the gas temperature.

$$\text{For } T = 290^\circ \text{C} \quad (i_e/i_+)_{\text{sat.}} = (m_+/m_e)^{\frac{1}{2}}(40V_e)^{\frac{1}{2}}, \quad (13)$$

$$\text{and} \quad \Delta V = \frac{1}{2}V_e \log\{1.58 \times 10^{14} m_+ V_e \mu^2 / \beta\}. \quad (14)$$

It must be remarked that $(i_e)_{\text{sat.}}$ cannot be measured by a plane probe at the higher pressures, as the drain of electrons from the region is too great to be replenished by a diffusive flow. However, it can be inferred from the electron temperature and the ion density found by other means.

The maximum value of d for which the high-pressure treatment is valid is a few per cent of the tube radius, since the tube wall is a probe taking a saturation ion current equal to the ionization in the positive column.

The low-pressure case

The determination of the potential distribution outside the sheath for the low-pressure case, when the ionization rate is constant, has been carried out by Langmuir (1929) in his study of a plasma bounded by two infinite planes. He obtains the potential shown in figure 1b.

The ordinates and abscissae are in terms of dimensionless parameters $\eta = V/V_e$ and s respectively. $S = 0$ corresponds to the centre of his discharge and is related to the rate of ionization and discharge dimensions. In making use of these results we have the following modifications. (i) Under our definition of the edge of the disturbed region, η , equal to between 0.1 and 0.01 (depending on the pressure), will mark the edge instead of $\eta = 0$. This has a negligible effect upon the result. (ii) The effective ionization rate is, in view of the finite size of the probe, augmented by ions entering from the side. (iii) Because the disturbed region is of the order of the probe dimensions or larger, edge effects are not small and the results are approximate. They are, however, not very sensitive to the shape of the potential distribution around the probe.

To determine the value of ΔV we have to find the value of η at which the sheath criterion is satisfied. The magnitude of the saturation positive-ion current is found from the values of the ion density at the sheath edge, and the mean positive-ion

velocity, obtained by graphical integration of the potential distribution in figure 1*b*. The positive-ion current to the probe at space potential will be somewhat larger than the mean normal random-ion current in the absence of the probe; larger because of the same considerations that led Langmuir to suggest his factor 2 mentioned already, though the actual magnitude of the factor will depend upon the anisotropy of the ion distribution.

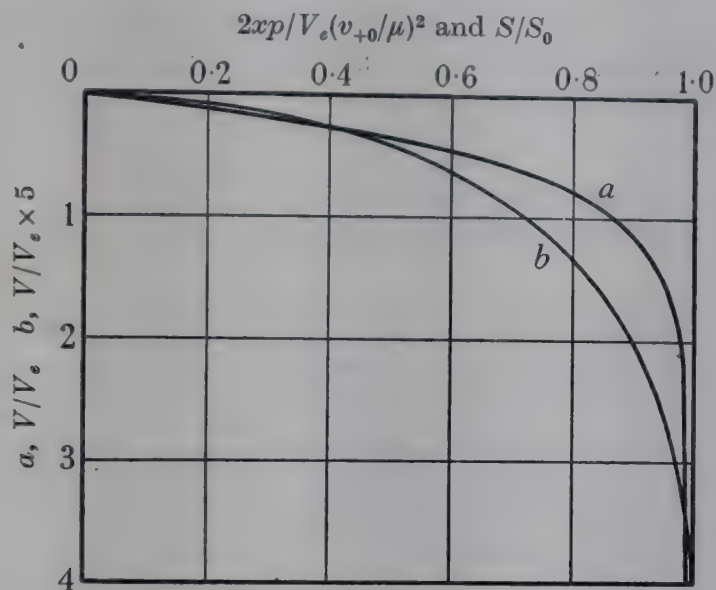


FIGURE 1. Potential distribution *a*, at high pressure, and *b*, at low pressure.

Graphical integration of the sheath criterion using the energy distribution of figure 1*b*, assuming various values of η for the sheath edge, reveals that the criterion is satisfied as soon as the upper integral becomes convergent, which occurs at $\eta = 0.7 \pm 0.1$.

$$\text{Thus} \quad \Delta V = 0.7V_e. \quad (15)$$

Graphical integration shows that the mean velocity of the ions at the sheath edge corresponds to $0.77(V_e)^{\frac{1}{2}}$ root volts. The ratio of the saturation currents is therefore (since $n_+^s = n_e^0 e^{-0.7}$)

$$\left(\frac{i_e}{i_+}\right)_{\text{sat.}} = n_e^0 \left(\frac{V_e}{2\pi m_e}\right)^{\frac{1}{2}} / n_+^s 0.77 \left(\frac{V_e}{m_+}\right)^{\frac{1}{2}} = 1.04 \left(\frac{m_+}{m_e}\right)^{\frac{1}{2}}. \quad (16)$$

TABLE 1. SUMMARY OF THEORETICAL RESULTS

	high pressure		low pressure
	$X_L/p > 0.75$	$X_L/p < 0.75$	
ΔV	$\frac{1}{2}V_e \log\left(\frac{V_e}{\beta} \frac{X_L}{p}\right)$	$\frac{1}{2}V_e \log(1.58 \times 10^{13} m_+ V_e \mu^2 / \beta)$	$0.7V_e$
$(i_e/i_+)_{\text{sat.}}$	$\frac{1.62 \times 10^7}{\mu} \left(\frac{V_e}{X_L/p}\right)^{\frac{1}{2}}$	$\left(\frac{m_+}{m_e} 40V_e\right)^{\frac{1}{2}}$	$1.04 \left(\frac{m_+}{m_e}\right)^{\frac{1}{2}}$

It will be shown later that the experimental results tend to confirm these conclusions. In the next section the apparatus used to make the experimental study is discussed.

3. THE APPARATUS

The discharge tube was of Pyrex glass 3 in. in diameter and 18 in. long. A nickel disk 2.5 in. in diameter formed the anode, while the cathode was a tungsten spiral carrying about 17 A. It was necessary to use a hot cathode at the lower pressures, and it was found convenient to use it at the higher pressures as well. The probe was introduced into the discharge through an orifice and a gland in a side arm, possible contamination of the gas in the discharge by the high-vacuum grease of the gland being prevented by maintaining the space between the gland and the orifice at a high vacuum. The tube was baked out at 300°C. Vapour traps were provided at all exits from the tube.

The current through the discharge was controlled by a saturated diode, the cathode heater current being adjusted to give an emission about 100 % in excess of the tube current.

During the measurements there was a copious flow of gas. Commercial argon (of 98.9 % purity) was used, from which any remaining oxygen had been removed by contact with hot copper. Oxygen, because of its ability to form negative ions, would invalidate the treatment. The only other impurity of importance in the argon was nitrogen, and the nature of the experiments made the presence of 1 % of this gas immaterial. Pressures were measured by a McLeod gauge, separated from the discharge tube by two liquid air traps with a tap (normally kept closed) interposed.

After preliminary experiments carried out with probes built from the grids of planar triodes, it became evident that the disturbance created in the plasma was often very great, and it was therefore decided to build as small a screened probe as seemed practicable. The technique employed was to transfer wires, stretched

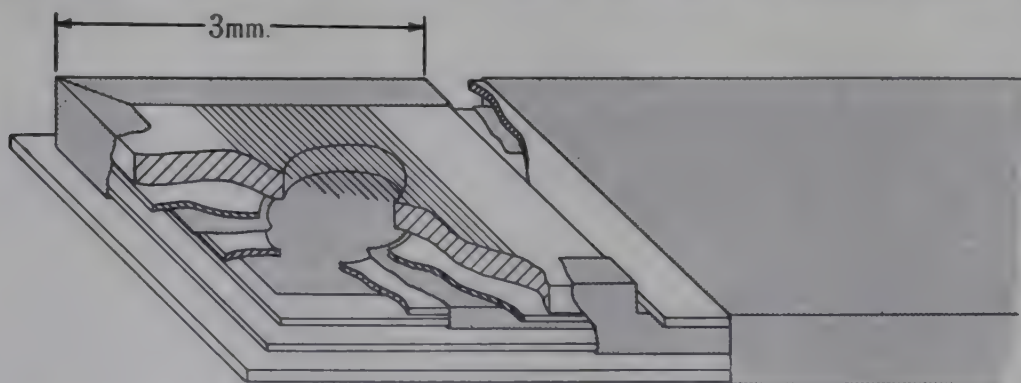


FIGURE 2. Construction of the probe. The platinum foil is shown stippled.

upon a larger frame, to a small one by brazing electrically in an atmosphere of hydrogen. By this method very small grids indeed can be made. The lower limit of the probe dimensions, however, is set by the mechanical and electrical problems of insulation. The grid frame was a 3 mm. square of molybdenum having a 1 mm. orifice crossed by 14 strands of 0.015 mm. diameter tungsten wire. Mica was used as an insulating medium between the grid and the platinum collector, but after short use it was noticed that undue heat owing to the outgassing process or to excessive electron bombardment reduced its insulation below the required figure of about $10^{12} \Omega$. The probe was therefore rebuilt with a system of guard-ring screening be-

tween grid and collector, and this, together with withdrawal of the probe from the centre of the tube during outgassing, completely eliminated the insulation trouble.

Mounting the probe on mica reduced the dimensions of the support to a minimum and reduced the disturbance of the discharge, especially when the probe was parallel to the electric field.

A d.c. amplifier requiring 0.1 V for full-scale deflexion was used for the collector current measurements.

The anode was biased by a variable stabilized voltage so as to make the probe floating potential the same as earth potential. This stabilization was important, as a drift here constitutes an error in the probe potential. Three readings were taken for each value of the probe potential:

(i) the total current to the grid and frame, corresponding approximately to the ordinary Langmuir probe characteristic;

(ii) the current with the collector biased +48 V to earth, giving the electron component;

(iii) the current with the collector biased -60 V to earth, giving the positive-ion component.

It was noticed that at the probe floating potential the electron component is always found to exceed the positive-ion component. This is due chiefly to secondary emission from the probe, which, while contributing to the floating potential equilibrium, will not contribute to the positive-ion component passing through the grid.

When it was large enough, the longitudinal field in the discharge was measured by finding the difference between the floating potential of the screened probe, when in the centre of the discharge, and that of a probe similarly situated 13 cm. nearer the anode. The latter probe was made of a nickel-copper alloy in place of the tungsten, molybdenum and silver-solder composition of the first. As a result of the difference in work function and shape the potential gradient measurements cannot be relied upon to a greater accuracy than about 0.05 V/cm.

4. THE SCREENING EFFECT OF THE GRID AND ASSOCIATED SOURCES OF ERROR

The grid-collector system departs from the ideal—a perfectly absorbing collector separated from the discharge by a potential barrier, variable in height—in four main ways.

(i) The transparency of the grid is not a constant but depends upon the ion and electron velocity and concentration distribution at its surface, and upon the electric field configuration around it. A constant transparency of 80 % has been assumed in this work.

(ii) The effective potential at the plane of the grid has an increment due to the collector potential. However, space-charge effects will greatly reduce the value 0.5 % of the collector potential given by Vogdes & Elder's formula (1924) for the effect in a charge-free space.

(iii) Ionization may occur in the space between the grid and collector, introducing a 'gas-amplification factor'. This factor may be estimated. It is given approximately by $(1 + p/10)$ and is negligible for those cases in which use has been made of the value of the electron current at space potential.

(iv) Secondary emission may take place at the grid or collector. Emission at the grid due to photons, metastable particles and positive ions appears in the characteristics as a residual electron current when a strong negative voltage is placed on the grid. It was never greater than 0.07 % of the electron random current. Secondary emission by electrons at the grid was also shown by experiment to be negligible.

Emission at the collector due to photons and metastable atoms is an important component but raises no difficulty, as it appears as a residual 'positive-ion' current when the grid is strongly positive. Emission due to action of the positive ions on the surface of the collector is only likely to be a small fraction of the collector current at the positive-ion energies involved. It would appear as a constant amplification factor, for the positive-ion current, and there is experimental evidence that such a factor if it exists is not far from unity.

Reflexion of positive ions or electrons at the collector would cause the recorded values of these currents to be too small. However, even though 20 % of the particles arriving were reflected, only a very small fraction would leave the collector at an angle sufficiently close to the vertical to enable them to escape its field, so that an accommodation factor of at least 95 % seems reasonable.

5. THE RESULTS OF EXPERIMENT

The primary data obtainable by the screened probe and not otherwise available from a Langmuir characteristic are threefold.

The positive-ion saturation current is obtained unaugmented by secondary, metastable or photo-emission. The form of the positive-ion curve in the neighbourhood of the space potential is revealed. The form of the electron characteristic is obtained directly instead of by subtraction of the extrapolated positive-ion current, and it can therefore be relied upon to higher retardation potentials.

The data shown in figure 3 *a*, *b* and *c* are typical for a range of pressures from 0.75 to 0.00085 mm. Hg in a discharge current of 100 mA, the probe being in the centre, edge-on to the cathode. Both the ion and electron curves show a residual current for high retardation voltages. That in the electron curve is due to secondary emission at the grid from various causes, and that in the positive-ion curve to emission from the collector by photons and metastable particles. Both the positive and negative residual currents show a tendency to increase as their respective retardation voltages on the grid are increased. In the electron curve this is accounted for by the intensified bombardment of the grid as it is made more negative. The rise at the positive end of the positive residual current curve is due to electrons, accelerated by the grid potential, ionizing and exciting in the region behind it.

(a) *The electron characteristics*

Figure 3 *c* gives a curve for the electron component of the current to the orifice, showing the details normally associated with a good Langmuir characteristic—the exponential rise, break-away at the space potential and the breakdown of the discharge to the probe. As the pressure is increased the break-away becomes less marked, until at 0.3 mm. Hg it has vanished altogether. The reason for this is apparent

from figure 4, which gives the semi-logarithmic plots of the electron currents. At the higher pressures the low diffusion rate of the electrons is inadequate to cope with the drain of electrons to the probe, with the result that the plasma density in the neighbourhood of the probe decreases, whilst its potential drifts with that of the probe giving distorted curves. The degree of drift of the space potential has been estimated by extrapolating the lower, linear part of the characteristic to the value of the electron current at space potential in the absence of the disturbance due to the probe. Since the latter quantity could not be determined directly at these pressures it was calculated from the electron concentration estimated, using Townsend & Bailey's (1922) mobility figures, from the longitudinal field strength, the tube current and the pressure. It has been marked on the curves by a cross.

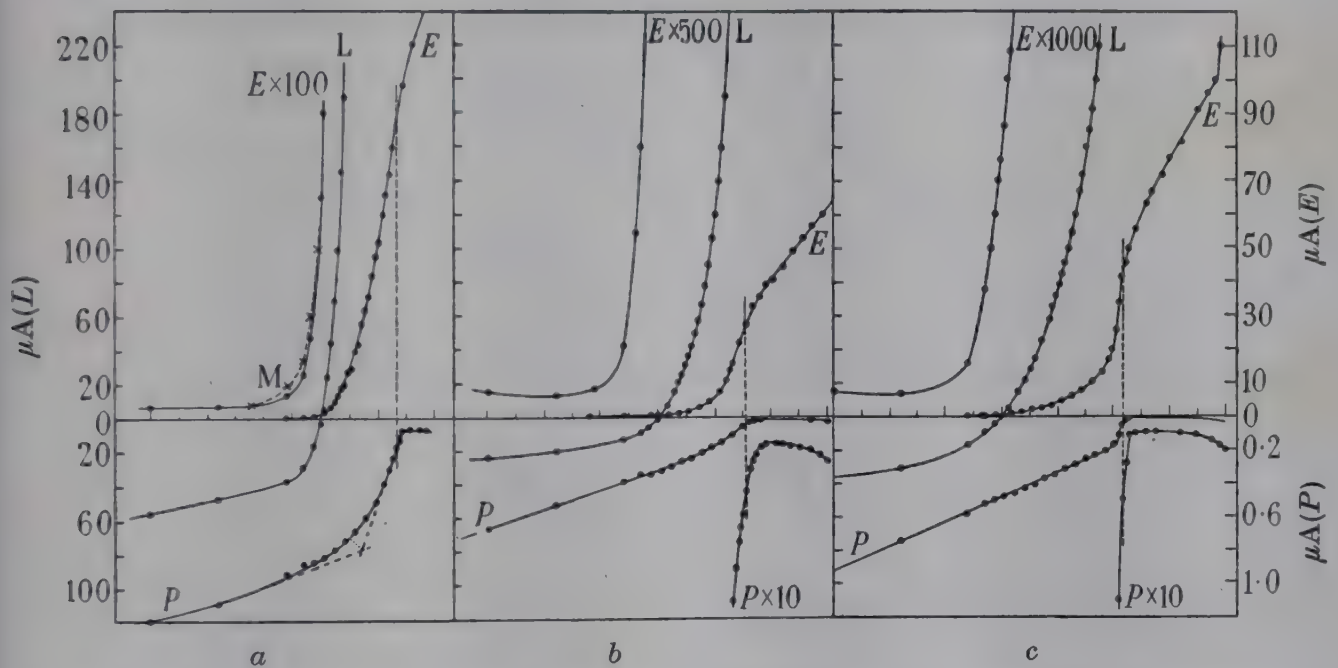


FIGURE 3. L =Langmuir curve; E =electron component; P =positive ion component; M =Maxwellian distribution. Pressures: a , 0.026; b , 0.004; c , 0.00085 mm. Hg. One division on the horizontal scale = 5 V.

Langmuir probe curves, taken with a 5 mm. diameter plane probe in argon at 0.1 mm. Hg pressure, show just the same effect of a drift in space potential, but curves for similar conditions in air show no drift, and are of conventional form. It is thought that an explanation of this can be found in the presence of negative ions in the air discharge, and it is hoped that the probing of electro-negative gases with the screened probe will elucidate this point.

At sufficiently low pressures the drift in the space potential is negligible. The curve at 0.004 mm. Hg shows an approximately linear semi-logarithmic plot with a well-defined break-away, corresponding to a roughly Maxwellian distribution of electron energies. The Langmuir characteristic for this case gave a temperature about 10 % higher and a 0.5 V difference in space potential. This agreement is probably as good as can be expected in view of the imperfect isotropy of the electron velocity distribution and the unsuitability of the probe frame as a Langmuir probe. It is unlikely that any useful information about the grid transparency can be

gleaned from the comparison. The ion densities determined from the two curves agree within the experimental limits of determining the ratio of the area of the probe to that of the orifice. At 0.00085 mm. Hg the electron curve is seen to depart very considerably from a true exponential (the effect is reflected in the Langmuir curve), showing a pronounced contribution from electrons of energies in the neighbourhood of 15 V. This contribution is not entirely absent at 0.004 mm. Hg, and its anisotropy is clearly demonstrated by the electron characteristics of figure 5 where the effect is seen to be strongest in the characteristic taken with the probe facing the cathode, suggesting that these electrons originate from this electrode.

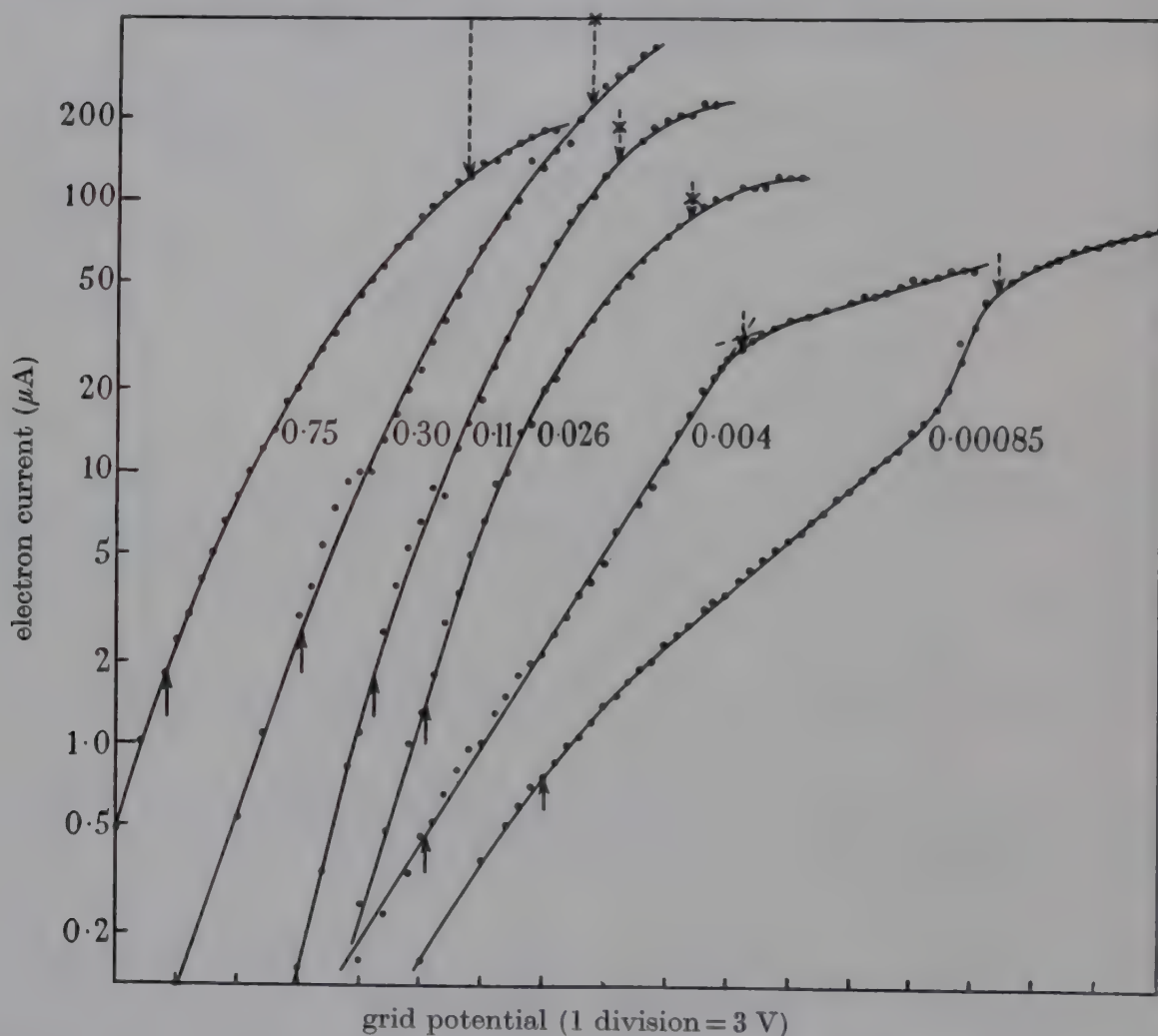


FIGURE 4. Semi-logarithmic curves for electron component. The upper arrows indicate space potentials and the lower arrows floating potentials. The numbers on the curves represent the pressures in mm. Hg.

If the residual electron current is subtracted from the recorded values to give the true electron current from the plasma, at high retarding potentials the resulting values are seen to be less than would be the case if the distribution were truly Maxwellian. The shape of the curve for a Maxwellian distribution has been shown dotted in figure 3*a*. This deviation is of considerable importance, since it occurs in the region where inelastic collisions are becoming important. (It is almost certainly caused by the drain of electron energy in this region by inelastic collisions and may well be largely due to the production of metastable argon.) The effect is not amenable

to quantitative treatment in the present experiments, owing to the presence of some nitrogen; but it serves as a warning against the extrapolation of an experimentally measured electron temperature to determine the rates of inelastic processes in the gas, and demonstrates the suitability of a screened probe for performing a Druyvesteyn analysis of the electron energies, at values of retarding voltage which would be too high to give a detectable electron current with an unscreened probe.

(b) *The positive-ion characteristics*

The typical features are most pronounced in the lowest pressure case, figure 3c. They are:

The positive-ion current at space potential is considerably below that at higher attracting voltages.

There is a saturation in the positive-ion current, which occurs, not at the space potential but at a voltage negative to this.

The ion current does not fall instantaneously to zero at the space potential, but remains finite for small retarding voltages.

The energy distribution of the normal component of the positive-ion velocities can be inferred from the retardation part of the characteristic, although the accuracy is not high owing to the small energy of the ions and the need to know the space potential with a high degree of precision. For this reason the only pressures at which an analysis of the positive-ion retardation curve could be performed with any degree of accuracy were in the region of 0.004 mm. Hg. At this pressure the value of X/p in the positive column is sufficient to give many of the ions an energy of the order of a volt; and 'temperatures' of this order were found. However, even in this region it must be remembered that the positive-ion velocity distribution is far from isotropic, and the experiments must be regarded as an indication rather than a precise measurement. The 'temperature' of the positive ions at 0.00085 mm. Hg pressure was less than 0.3 V, but in both this and the higher pressure case the distribution was very deficient in the high-energy components.

The shapes of the positive-ion curves at the higher pressures are all very similar, and it is noticeable that in spite of the decrease in density of ionization with decreasing pressure, the value of the saturation positive-ion current remains practically the same over the pressure range from 0.75 to 0.026 mm. Hg. At constant pressure, however (0.3 mm. Hg), the saturation currents are accurately in the ratio of the tube currents. For a ratio 1 : 0.56 : 0.17 in the latter the ratio of the former are 1 : 0.57 : 0.16.

It is not easy to decide the precise voltage at which saturation begins (i.e. a sheath forms), but for the higher pressures it has been taken as the point of intersection of the extrapolated rising part of the curve with the extrapolated saturation part (shown dotted in figure 3a). For the lower pressures it has been estimated from the semi-logarithmic curves. In the case of the distorted curves the true space potential has been found (for the purpose of estimating ΔV) by extrapolation of the linear part of the electron curves, as already described.

The data of experiments with the probe edge-on to the cathode are summarized in table 2. The current was near 100 mA in each case except in two experiments at

TABLE 2. SUMMARY OF RESULTS WITH PROBE IN CENTRE EDGE-ON TO CATHODE

pressure (mm. Hg)	tube current (mA)	X_L/p (V/mm. Hg)	electron temperature $V_e = KT_e/e$	saturation currents		ionization density ($\times 10^{-8}$ cm. ³)	residual currents		measured ΔV (V)
				electron (μ A)	positive ion (μ A)		negative (μ A)	positive (μ A)	
0.75*	103	0.97	1.8	(1.750)	0.58	700	52	83	4.0
0.30*	103	2.8	1.7	(460)	0.60	200	72	80	5.0
0.30*	58	3.6	1.6	(140)	0.34	71	40	44	4.0
0.30*	18	5.7	1.5	(30)	0.097	14	11	19	4.0
0.113	100	6.8	1.7	(180)	0.60	75	36	46	2.5
0.026	98	16.5	1.9	(100)	0.60	40	36	64	0.5
0.023*	100	10.4	2.7	(150)	0.62	53	30	75	4.5
0.0047	100	—	3.7	24	0.17	7.0	10	15	2.5
0.0043	100	—	3.9	24	0.12	8.1	9	13	2.5
0.0040	96	[100]	3.9	27	0.11	8.1	15	16	3.0
0.00095	94	—	[4.4]	38	0.14	10.7	11	8	2.0
0.00085*	118	—	[5.4]	50	0.20	12.8	7	9	2.5

Results marked with an asterisk were taken before the full outgassing procedure was adopted.

Results given in square brackets are doubtful.

Results given in round brackets are calculated from the other data, not measured.

0.3 mm. Hg pressure, in which it was 58 and 18 mA respectively. The shape of the electron and positive-ion curves did not show any marked dependence on the tube current, except that the onset of positive-ion saturation appeared to be more sharply defined at the lower tube current. The current referred to as the positive-ion saturation current was found by dropping a perpendicular to the curve from the intersection described above, and subtracting the residual positive-ion current. It must be emphasized that these constructions are highly empirical, but some construction is essential if any comparison is to be made between theory and experiment. The six experiments marked in the table with an asterisk were carried out without the full outgassing procedure, as at that time there were no facilities for baking the tube other than a gas flame.

TABLE 3. COMPARISON OF THEORY AND EXPERIMENT FOR PROBE IN CENTRE EDGE-ON TO CATHODE

pressure (mm.Hg)	ΔV (V)			$(i_e/i_+)_{\text{sat.}}$		
	measured	calculated		measured	calculated	
		high pressure	low pressure		high pressure	low pressure
0.75	4.0	4.4	1.3	3,000	2,000	280
0.3	5.0	3.3	1.2	770	1,100	280
0.3	4.0	2.8	1.1	410	970	280
0.3	4.0	2.3	1.0	310	740	280
0.113	2.5	2.5	1.2	300	730	280
0.026	0.5	2.1	1.3	170	490	280
0.023	4.5	4.0	1.9	240	740	280
0.0047	2.5	—	2.6	140	—	280
0.0043	2.5	—	2.7	200	—	280
0.0040	3.0	(2.1)	2.7	250	(290)	280
0.00095	2.0	—	(3.1)	270	—	280
0.00085	2.5	—	(3.8)	250	—	280

Using the results of table 2 and the formulae tabulated in table 1 we can make a comparison between theory and experiment, bearing in mind, however, that the edge effects of the probe are not negligible and the results cannot be expected to do more than indicate the trend. This has been done in table 3. Unfortunately the ion density was not really sufficient to make the high pressure theory applicable (to be valid it can be shown that the ion density should exceed $10^8/\lambda^2$). Failure to fulfil this condition means that the sheath extends into the diffusive flow region. This would not be expected greatly to affect the calculated saturation current ratio, but it means that neither the high nor low pressure theoretical values of ΔV are applicable to the high pressure experiments. However the comparison shows that up to a pressure of 0.1 mm.Hg the low pressure formulae both for ΔV and the saturation current ratio are in reasonable agreement, while at the higher pressures the saturation current ratio tends to the much higher value given by the high pressure theory.

TABLE 4. SUMMARY OF RESULTS FOR VARIOUS PROBE POSITIONS AND ORIENTATIONS

pressure (mm.Hg)	tube current (mA)	dis- tance from centre (cm.)	orientation	electron tem- perature V_e (V)	saturation current		ion density ($\times 10^{-8}$ / cm. ³)	residual currents		measured ΔV (V)		random current ratio based on			saturation current ratio
					electron (μ A)	positive ion (μ A)		negative (μ A)	positive (μ A)			tem- perature	S.P.	mean	
0.0040	96	0	side	3.9	27	0.11	8.1	15	16	3.0	3.0	500	900	700	250
0.0043	104	0	side	3.9	24	0.12	7.2	9	13	2.5	2.5	420	590	510	200
0.0047	100	0	side	3.7	24	0.17	7.4	10	15	2.5	2.5	430	300	370	140
0.0045	102	1	side	3.7	24	0.13	7.4	9	13	2.5	2.5	440	440	440	180
0.0047	100	2	side	3.7	20	0.11	6.2	8	10	2.5	2.5	500	500	500	180
0.0049	101	3	side	4.2	15	0.064	4.4	5	6.6	2.5	2.5	660	880	770	230
[0.0055]	[94]	3	side	3.7	16	0.074	5.0	5	7.8	3.5	3.5	780	890	840	230
0.0048	100	0	facing cathode	4.4	39	0.24	11.0	12	20	2.5	2.5	460	270	370	160
0.0038	97	0	facing cathode	4.7	34	0.18	9.3	14	14	3.0	3.0	440	470	450	190
0.0046	99	0	facing anode	3.9	18	0.17	5.4	10	20	0	0	—	110	—	110
0.0035	96	0	facing anode	3.3	14	0.11	4.6	11	12	0	0	—	130	—	130

(c) *The effects of the probe orientation and position*

Table 4 summarizes the data obtained with the probe in discharges of about 100 mA at pressures in the region of 0.004 mm. Hg for a number of different probe orientations and positions. For each of these experiments semi-logarithmic plots of the electron and positive-ion components have been made, and the current ratio at the space potential found. Three of these graphs are shown in figure 5*a*, *b* and *c*. Figure 5*b* is typical of all the results obtained at these pressures, with the probe edge-on to the cathode, irrespective of the radial position of the probe. The value obtained for the current ratio at space potential is very sensitive to the position of the space potential, owing to the high slope of the positive-ion curve, and for this reason an alternative value has been found, on the basis of the ratio of the apparent positive-ion temperature to that of the electrons. Where appropriate the average of these two results has been taken in making comparisons with theory.

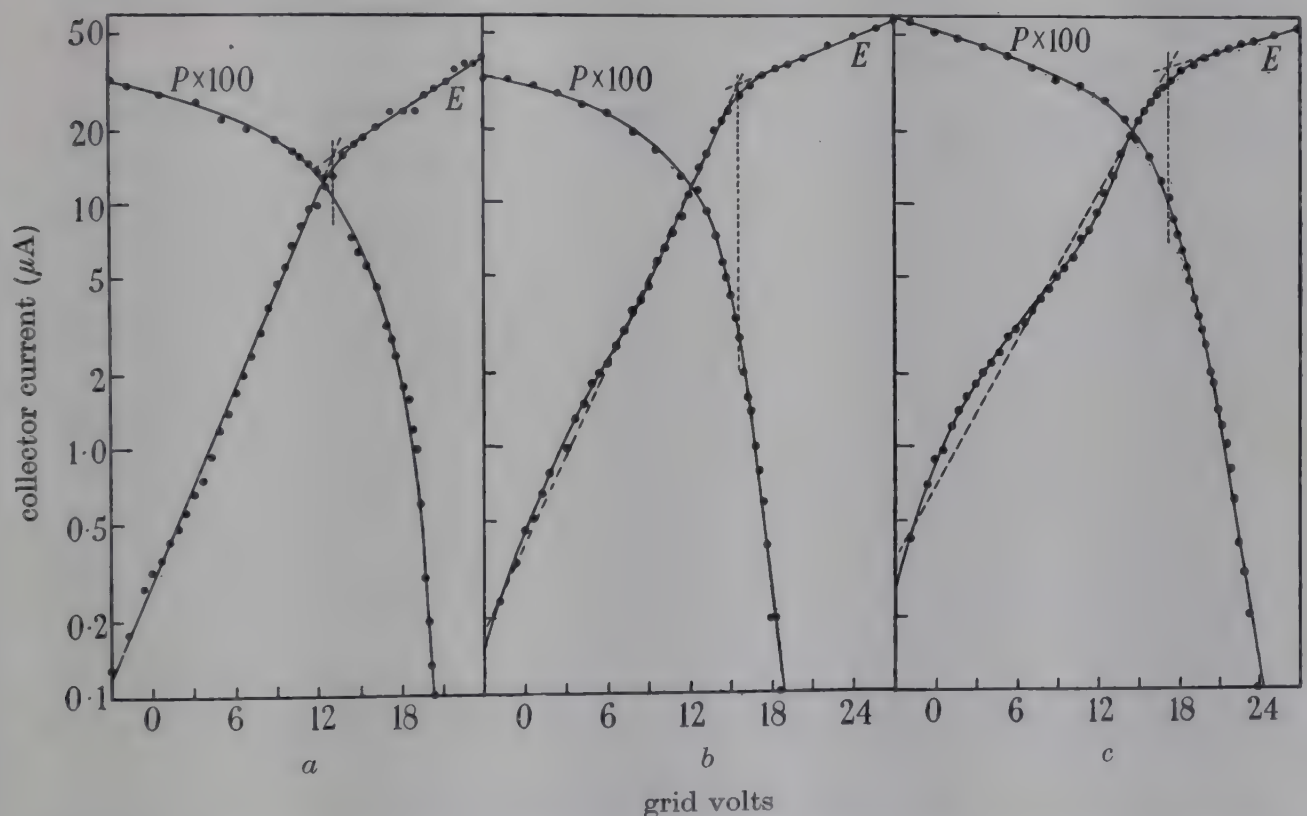


FIGURE 5. Semi-logarithmic curves for positive ions and electrons. *a*, probe facing anode; *b*, probe side-on to cathode; *c*, probe facing cathode.

Referring first to figure 5*b* positive-ion saturation is seen to start well negative of the space potential. Figure 5*c*, which was obtained with the probe facing the cathode, is similar in general form, though the much-increased electron current, together with the departure of the electron curve from linearity, indicates a marked anisotropy of the electron-velocity distribution. The roughly proportional increase in the positive-ion current at space potential may be attributed to the redistribution of potential on the cathode side of the probe, which must occur on the introduction of the probe, in order to maintain the space charge neutralization otherwise disturbed by the shielding of this region from the main flow of positive ions. The curve

obtained with the probe facing the anode (figure 5a), however, is quite different. In virtue of the longitudinal field the positive-ion velocities satisfy or very nearly satisfy the sheath criterion without the probe being made negative to the space. ΔV is therefore effectively zero and saturation occurs close to the space potential. For the same reason the positive-ion current is greater in spite of a decrease in electron current due to the sheltering of the region in front of the probe from the electron stream. Mean values for two experiments at each orientation are given in table 5,

TABLE 5. EFFECT OF PROBE POSITION AND ORIENTATION

orientation position	facing cathode centre	facing anode centre	side on			
			centre	1 cm. out	2 cm. out	3 cm. out
ion density $\times 10^{-8}$	10.2	5.0	7.3	7.4	6.2	4.4
$i_e \mu A$	36	16	24	24	20	16
$(i_e/i_+)_{s.p.}$	410	120	440	440	500	770
$(i_e/i_+)_{sat.}$	175	120	170	180	180	230

together with the results of a series of experiments with the probe edge-on at various distances from the centre of the tube, taken on the same occasion without extinguishing the arc. There seems to be a slight tendency for the saturation current ratio to increase with increase in the radial distance of the probe from the centre, and a very definite tendency for the space potential current ratio to do so. (These tendencies are confirmed by other similar readings.) At first sight this seems strange, since the energies of the positive ions must be greater at the edge of the plasma than in the centre. It would appear, however, to be due to the high ratio of drift to random positive-ion current, so that the ion current to the probe is governed largely by the extent of the volume in front of the probe from which ions may be scattered towards it, the potential distribution in this volume and the solid angle subtended by the orifice at various regions in the volume. The volume of ionization from which ions can be scattered into the orifice is clearly much less at the edge of the tube than in the centre. The construction of the probe used in the present experiments unfortunately prevented it being orientated normal to the tube radius.

(d) *Secondary emission from the collector*

One of the chief reasons for uncertainty in the true value of the positive-ion current to a Langmuir probe is the unknown contribution from emission by photons and metastable particles (see Uyterhoevn & Harrington 1930). Quite a large amount of information about this can be gained from these experiments, though without a programme designed for the purpose the results are empirical, and it is not possible to say definitely whether the effect is principally due to metastable atoms or photons. It seems likely, however, that the metastable atoms are chiefly responsible since the increase in emission on turning the probe to face up or down the column is only 40 % while the depth of plasma in front of the probe is increased fivefold. This was done at 0.004 mm. Hg where the gas kinetic mean free path is about 1.5 cm. (0.2 of the tube diameter). The argument, however, would not be valid if the diffusion cross-section for the active photons was of the order of the gas kinetic cross-section, although

TABLE 6. VARIATION OF SECONDARY EMISSION WITH PRESSURE, CURRENT, PROBE ORIENTATION AND POSITION

Tube current $\approx$ 100 mA						
pressure (mm.Hg)	0.75	0.3	0.11	0.025	0.004	0.0009
secondary emission (%)	14	13	8	12	13	5
pressure = 0.3 mm. Hg						
current (mA)	103	58	18			
secondary emission (%)	13	12	20			
tube current = 100 mA; pressure = 0.004 mm. Hg						
orientation	facing cathode		facing side		facing anode	
secondary emission (%)	11		12		8	
tube current = 100 mA; pressure = 0.004 mm. Hg						
position (cm. from centre)	0	1	2	3		
secondary emission (%)	12	10	9	11		

this is thought unlikely. Table 6 gives the variation of the secondary emission as a percentage of the positive-ion saturation current, with pressure, current, orientation and position. The data are mean values, taken from tables 2 and 4.

In conclusion, the author wishes to thank Dr E. H. S. Burhop for his close interest and very valuable assistance, Professor H. S. W. Massey, F.R.S., and Mr D. R. Bates for many helpful discussions, and Mr E. Aström for sparing time from his own research to help in the experiments. Thanks are also due to the Department of Scientific and Industrial Research for financing the project.

REFERENCES

- Druyvesteyn, M. J. 1930 *Z. Phys.* **64**, 781.
 Emeléus, K. G., Greeves, F. D. & Montgomery, E. 1936 *Proc. R. Irish Acad. A*, **43**, 35.
 Frisch, S. & Kagan, J. 1943 *J. Phys. U.S.S.R.* **7**, 208.
 Greeves, F. D. & Johnston, J. E. McF. 1936 *Phil. Mag.* **21**, 659.
 Langmuir, I. 1929 *Phys. Rev.* **33**, 950.
 Langmuir, I. & Mott-Smith, H. 1924 *Gen. Elect. Rev.* **27**, 449, 538, 616, 762.
 Langmuir, I. & Tonks, L. 1929 *Phys. Rev.* **34**, 876.
 Sena, L. 1935 *Trans. All. Un. Electrotech. Ass.* **5**, 42.
 Sena, L. 1946 *J. Phys. U.S.S.R.* **10**, 179.
 Tonks, L., Mott-Smith, H. & Langmuir, I. 1926 *Phys. Rev.* **28**, 104.
 Townsend, J. S. & Bailey, V. A. 1922 *Phil. Mag.* **44**, 1033.
 Uyterhoevn, W. V. & Harrington, M. C. 1930 *Phys. Rev.* **36**, 709.
 Vodges, F. B. & Elder, F. R. 1924 *Phys. Rev.* **24**, 683.

The inelastic scattering of protons by magnesium, aluminium and other light elements

BY E. H. RHODERICK, *Cavendish Laboratory, University of Cambridge*

(Communicated by P. I. Dee, F.R.S.—Received 6 October 1949)

The energy spectrum of 4.5 MeV protons scattered by some light elements was determined by using a proportional counter in conjunction with absorbing foils. Magnesium gave groups attributed to levels at 1.36 ± 0.03 and 1.82 ± 0.04 MeV respectively above the ground state, the former being almost certainly associated with ^{24}Mg . The angular distribution and excitation function of this group were examined in some detail. The proton groups from aluminium corresponded to levels at about 1 and 2.15 ± 0.05 MeV respectively above the ground state. The lower level has a complex structure and probably consists of two levels at about 0.80 and 0.97 ± 0.02 MeV, the amplitudes of the corresponding proton groups showing a marked angular dependence. No inelastically scattered protons were observed from carbon, oxygen, or copper, while beryllium gave a group consistent with the existence of a metastable state in ^9Be at 2.39 ± 0.05 MeV above the ground state. A short theoretical discussion of the mechanism of the process is given which suggests that inelastic scattering is predominantly a (p, p) process involving formation of a compound nucleus.

INTRODUCTION

The inelastic scattering of protons has been observed by several workers (Wilkins 1941; Powell, May, Chadwick & Pickavance 1940; Dicke & Marshall 1943; Davis & Hafner 1948*a*; Fulbright & Bush 1948), and has proved a valuable method for the determination of nuclear energy levels. The basic process is illustrated in figure 1*a*, the net result being that the scattering nucleus is raised to an excited state at the expense of the kinetic energy of the system, so that observation of the energy spectrum of the scattered protons gives information about the nuclear levels involved.

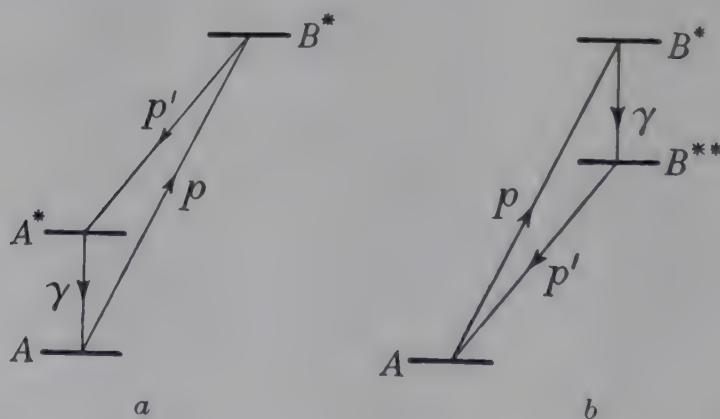


FIGURE 1. Schematic representation of scattering process.
A = scattering nucleus, *B* = compound nucleus.

Before one can interpret the results, it is necessary to exclude the possibility of a process such as that shown in figure 1*b*, in which the γ -ray transition takes place not in the residual but in the compound nucleus. Such a process would give rise to inelastically scattered protons, but would not involve any excited states of the

residual nucleus. In the experiments to be described later, the compound nucleus B has an excitation energy of about 10 MeV, while the loss in energy of the scattered proton is of the order of 1 MeV. For γ -ray transitions of multipole order l , the transition probability is proportional to E^{2l+1} , where E is the γ -ray energy, so that the probability of a process such as that shown in figure 1*b* should be at most 10^{-3} times as great as that of a direct transition to the ground state of B . Such a process should therefore be much less probable than simple capture, which in turn is far less likely than the emission of a proton because of the smallness of the radiative widths of nuclear levels. We may therefore assume that inelastic scattering follows the scheme of figure 1*a*, and that the loss of energy of the proton, measured in the co-ordinate system in which the centre of mass is at rest, is equal to the excitation energy of the residual nucleus.

The remainder of this paper describes some experiments on the inelastic scattering of protons which have been carried out with the aid of the Cambridge cyclotron. At the time of the experiments the cyclotron was producing protons of 4.5 MeV energy, which sets a limit to the number of levels excitable by proton bombardment. For this reason, the object of the work was not simply to investigate nuclear energy levels, but also to investigate the scattering process in some detail, with a view to elucidating its mechanism. Most of this detailed work was carried out on magnesium, and will be described in full. The account of the investigations of other elements will be confined to points of particular interest in each case. A preliminary account of the work has already appeared (Rhoderick 1949).

APPARATUS

Basically, the method adopted in these experiments resembles that of Dicke & Marshall, the only significant difference being that provision was made for observation at several angles. The spectrum of the scattered protons was investigated by using absorbing foils in conjunction with a proportional counter. The construction of the vacuum chamber is shown in figure 2. The least practicable separation of the windows was 20° , and these were placed at even multiples of 10° on one side, and at odd multiples on the other, so that the scattered protons could be observed at 10° intervals from 20 to 160° . The windows were of mica, of stopping power equivalent to about 1 cm. of air, the area of each window being determined by a $\frac{1}{8}$ in. diameter hole in the brass ferrule on which it was mounted. The lid of the chamber contained a Wilson seal in which the target holder was mounted so that it could be rotated and also raised out of the path of the proton beam. The incident beam, which was effectively parallel on reaching the scattering chamber, was defined by a stop having a $\frac{1}{8}$ in. diameter hole. Before reaching the target, the beam passed through two further stops with $\frac{3}{16}$ in. diameter holes which served to block protons scattered from the edge of the first stop. The geometry was such that protons passing through a given window were scattered through an angle within $\pm 1.8^\circ$ of the mean angle.

The counters were standard laboratory β -counters operated in the proportional region, the filling consisting of a 10 : 1 mixture of pure argon and alcohol at a total pressure of 10 cm. Several different types of window were used; all were of mica,

with stopping powers varying from 2 to 4 cm. of air. The thinnest windows either had very small apertures or were mounted on grids. The mechanical support for the counter fitted over a lip on the window mount, which located it accurately relative to the chamber window, the counter being slightly offset to avoid the possibility of protons hitting the glass bead on the end of the anode. The counter used for measuring angular distributions was fitted with an iris which eliminated any protons scattered from the edge of the chamber window. The anode of the counter was connected to the h.t. supply through a resistance of 150,000 ohms; this low value enabled the charge on the wire to leak away very quickly, so producing a pulse of short duration with a very rapid rise. The counter was connected to a 'fast' feedback amplifier having a flat response from about 10 keyc. up to 5 meyc. The

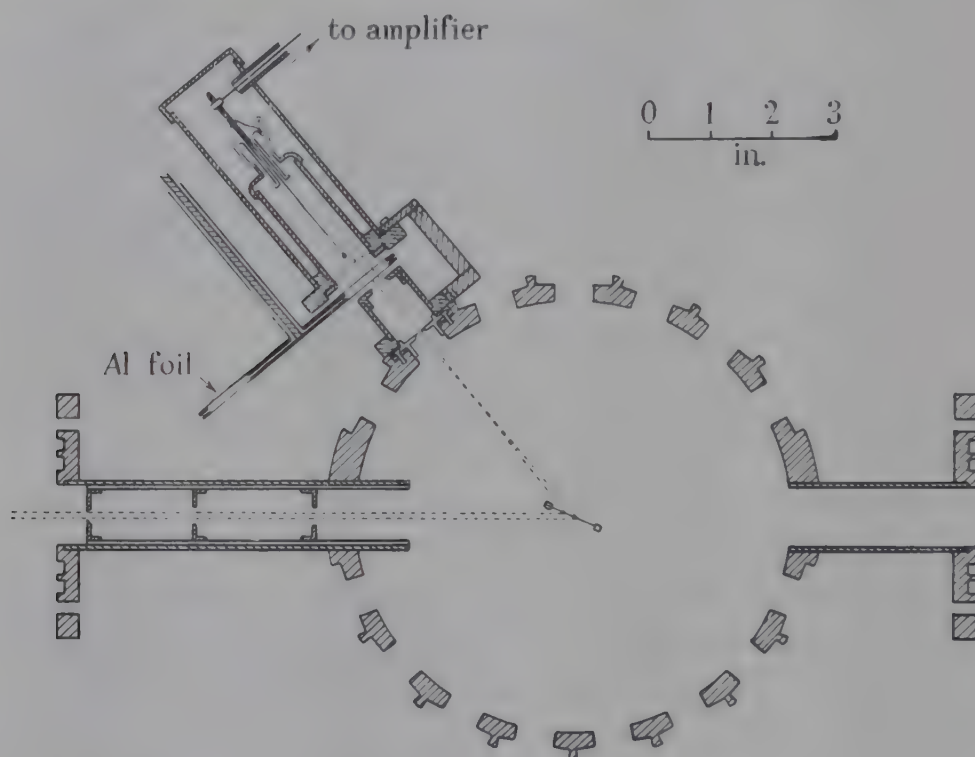


FIGURE 2. Scattering chamber and counter assembly, showing one window only.

counters were operated with a gas amplification of about 50, under which conditions the signal-to-noise ratio produced by a proton at the peak of the Bragg curve was about 20 : 1. The output stage of the amplifier was connected to a discriminator and thence to a scale-of-100 recorder. The overall resolving time of the system was found to be about $5\mu\text{sec}$. The main amplifier and recording equipment were housed in a screened cabin some distance from the scattering chamber, a preamplifier being used to minimize the length of lead attached to the counter. To enable a second counter to be used for monitoring purposes, the amplifier and recording system were duplicated.

For speed of operation, it was desirable that the absorbing foils be changeable by remote control from the cabin. It appeared that the simplest and most reliable way of doing this was to use Post Office 'preselectors', which are power-driven rotary switches. These were modified so as to enable them to be set in any of six equally

spaced positions. Two preselectors were used to rotate two concentric disks 4 in. in diameter, having six $\frac{7}{8}$ in. diameter holes equally spaced around their circumferences, over which were mounted the absorbing foils. A total of thirty-six combinations could thus be chosen at will. The absorbing foils consisted of aluminium of various thicknesses, punched out of sheets with a steel punch whose diameter was accurately known. The weight of each foil was measured to 0.1 mg., and the uniformity of the foils checked by comparing the weights of several disks cut from the same sheet. Suitable combinations of these foils were cemented over the holes in the foil changers, the combinations being so chosen that the total thickness of aluminium could be varied in steps of approximately 1.5 mg./cm.², which is equivalent to 1 cm. of air. Intermediate observations could be made by interposing an additional foil of thickness 0.8 mg./cm.² For some purposes it was preferable to use an air cell rather than aluminium foils; the one which was used had a total absorption path of 23 cm., and was connected to the scattering chamber and to the counter with rubber gaskets.

In carrying out these experiments, use was made of a magnet which had been constructed for bringing the cyclotron beam to a focus at a point some distance from the cyclotron tank (Rhoderick 1948). This was used to produce a parallel beam rather than a convergent one, and enabled the scattering chamber to be situated at a distance of 15 ft. from the cyclotron, with a water tank 3 ft. thick intervening. With this arrangement, the background measured with a proportional counter was negligible. In order to measure the energy of the incident beam, a simple magnetic analyzer of the type described by Creutz & Wilson (1946) was constructed. This was attached to the scattering chamber at the end remote from the cyclotron, the beam passing into it when the target holder was raised. The experimental lay-out is shown in figure 3. The various parts of the apparatus were lined up in succession with the aid of fluorescent screens.

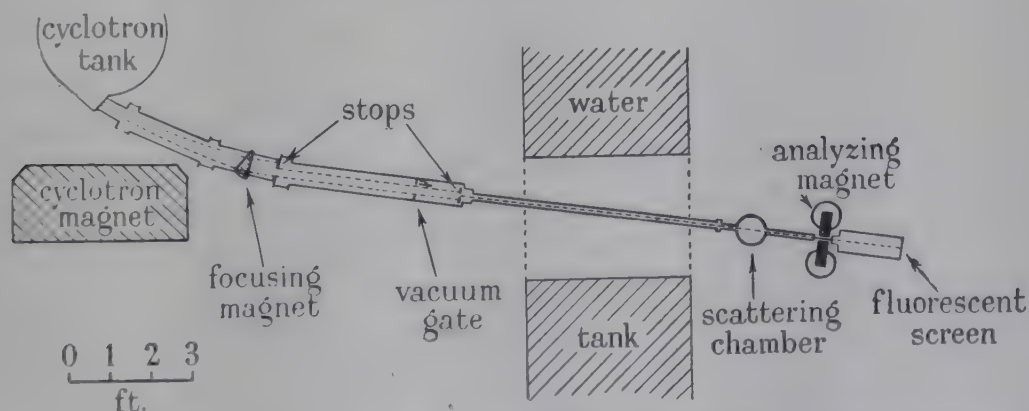


FIGURE 3. General assembly of scattering chamber and focusing magnet.

PROCEDURE

To determine the energy spectrum of the scattered protons, the counter was made to give a differential range curve by setting the discriminator bias at a fairly high value, so that only those protons passing through the counter near the end of their range were recorded. This procedure is sometimes referred to as 'peaking'. A

thorough analysis of the effect of discriminator bias on the width of the peaks showed that the optimum bias setting is about 0.7 times the pulse height produced by a proton at the maximum of the Bragg curve; increasing the bias beyond this value decreases the counting rate without any significant decrease in the width of the peak. The beam was monitored by using a second counter operated with very low bias and placed at such an angle that the counting rate was sufficient to make statistical errors in the monitor negligible. Counts were normally taken at intervals of about 1.5 mg./cm.², but in the vicinity of a peak intermediate points were obtained using the 0.8 mg./cm.² foil. The position of the scattering foil during these measurements is very important. Dicke & Marshall adopted the arrangement shown in figure 4*a*, with the normal to the foil bisecting the angle between the incident and scattered protons. It is easily seen from the diagram that, since the

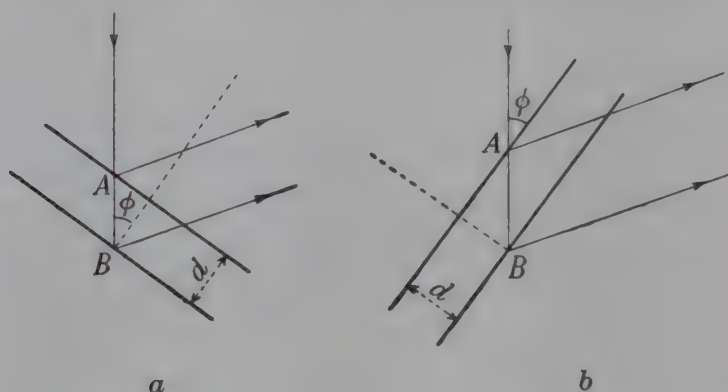


FIGURE 4. Effect of target disposition on proton paths.

incident proton may be scattered anywhere in the foil, the total path length of the proton inside the scatterer varies between zero and $2 \times AB$. The scattered protons will therefore have a spread in energy equal to the energy lost in traversing a distance $2d \sec \phi$ in the scatterer. However, if the foil is placed as in figure 4*b*, so as to bisect the angle between the incident and scattered protons, the path length of any proton inside the scatterer is independent of the position of the collision, and is equal to AB , or $d \operatorname{cosec} \phi$. In this case there will be very little spread in energy of the scattered protons due to the thickness of the scattering foil. (There will, in fact, be a small effect even in the latter case, because the proton loses energy in the collision, and the range-energy relation is not linear.) It is significant that whereas the range curves given by Dicke & Marshall have a half-width of 8.5 mg./cm.² of aluminium, the half-width obtained in the experiments described here was only 2.5 mg./cm.², the difference being attributable to the change in target geometry. Under these conditions, the resolving power for groups of equal amplitude was about 250 keV. This figure is about 40 % greater than the lower limit imposed by straggling, and is very near to the best obtainable with this type of apparatus.

To calculate the Q values corresponding to the inelastic groups, the following procedure was adopted. It is difficult to measure absolute ranges with a proportional counter, as the relationship between the mean range of the group and the position of the maximum of the counts-range curve is a complicated one, depending on the depth of the counter and the bias voltage. However, the difference in range

of two groups is given exactly by the distance between the corresponding maxima, without the addition of any corrections. Now the absolute range of the elastic group can be obtained from the range-energy curves of Livingston & Bethe (1937) if the energy of the incident protons is accurately known, so that measurement of the beam energy together with the distance between the elastic and inelastic peaks suffices to determine the absolute range of the inelastic group and hence its energy. From this, the Q can be evaluated in the usual manner. The ranges in aluminium were converted to ranges in air by following the procedure of Livingston & Bethe.

It is worthy of note that the proton energies measured with the magnetic analyzer were consistently lower than the values calculated from the known frequency of the r.f. voltage and the distance of the exit slit from the centre of the dees, sometimes by as much as 8 %, which is well outside the error of the measurement. When it is realized, however, that a variation of 8 % in the energy corresponds to a change of 4 % in the radius of the final orbit, and that the value of the latter was 38 cm., it will be seen that it has to be in error only by 1.5 cm. to explain the discrepancy. The radius of the final orbit is undoubtedly subject to an uncertainty of this order, and this explanation is supported by the observation that when the cyclotron ion source was replaced there was a noticeable change in energy, presumably because it was not put back in precisely the same position. The slit in the analyzer was horizontal, and the deflexion vertical; because of the inhomogeneity in energy of the incident protons, the line produced by the beam on the fluorescent screen underwent a pronounced tilt when it was deflected, the direction of tilt being such that the most energetic component was on the outside (i.e. farthest from the centre of the dees). The magnitude of the tilt indicated that the total spread in energy in that portion of the beam which passed through the $\frac{1}{8}$ in. diameter defining stop amounted to about 1.5 % of the mean energy, or 60 keV.

EXPERIMENTAL RESULTS

(a) *Magnesium*

The magnesium target consisted of a thin foil obtained by rolling magnesium ribbon to a thickness of 0.0004 in. (1.8 mg./cm.²). The spectra of the scattered protons obtained at different angles are shown in figure 5, the incident proton energy being 4.73 MeV. At each angle the peak having the greatest range consists of elastically scattered protons, while the inelastically scattered protons constitute the shorter range peaks. It will be observed that the range of the elastic peak decreases with increasing scattering angle. There are two reasons for this: first, the effect of the recoil of the scattering nucleus is greater at backward angles, and second, the effective thickness of the target, which is $d \operatorname{cosec} \phi$, increases as ϕ decreases. The experimental arrangement was such that only inelastically scattered protons corresponding to an excitation energy under 2.5 MeV could penetrate the counter and chamber windows. The results of Dicke & Marshall, obtained at a scattering angle of 135° with a proton energy of 6.9 MeV, gave inelastic peaks corresponding to excitation energies of 1.32, 2.74, and 3.88 MeV respectively. Only the lowest of these would therefore be detectable under the conditions described here.

With the improved resolution of the present experiment it is apparent that the first inelastic peak is accompanied by a smaller peak on the side corresponding to higher excitation energy. (By 'first inelastic peak' is meant the peak nearest to the elastic peak. This group corresponds to the first excited state of the target nucleus.) The Q values for the first inelastic group, calculated from the absorber thickness in the manner described previously, lead to a mean excitation energy of 1.45 MeV for the level of the residual nucleus involved. This Q value remained constant with angle to within ± 0.015 MeV, which represents the relative error in the measurements at different angles. The absolute value of the Q , however, is more uncertain,

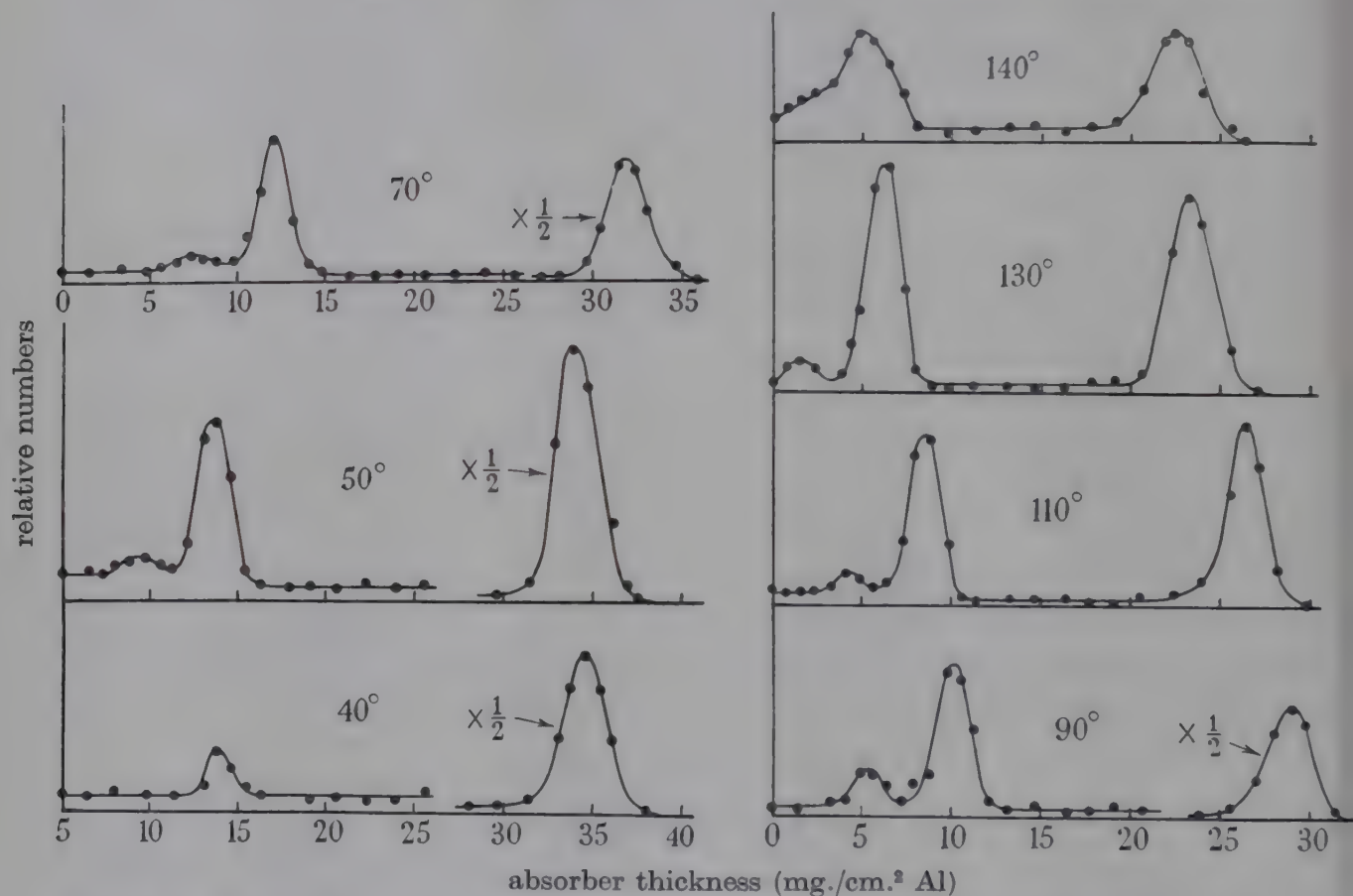


FIGURE 5. Spectra of protons scattered by magnesium ($E_0 = 4.73$ MeV). The vertical scale is not necessarily the same at all angles.

since it depends on the stopping power of the absorbing foils. The result was checked by using the air absorption cell instead of the aluminium foils, and a significantly lower value obtained, namely, 1.36 MeV. In view of the large amplitude of the group in question, it is almost certainly associated with the predominant isotope ^{24}Mg . The level involved is probably that responsible for the 1.38 MeV γ -ray emitted following the β -decay of ^{24}Na , which has been measured with great accuracy by Siegbahn (1946). Since the error in a single determination with the air cell (due to uncertainties in the exact position of the peaks, and in the beam energy) is estimated as ± 0.03 MeV, the result obtained with the air cell agrees with the γ -ray measurements within the experimental error. We shall therefore assume that the measurements with the air cell were correct, and that the energy of the first excited state of ^{24}Mg , as determined by inelastic scattering, is 1.36 ± 0.03 MeV

above the ground state. The discrepancy involved in using the aluminium foils is mysterious, and was never resolved owing to lack of opportunity for accurate calibration; because of it, absolute determinations of Q values were made with the air cell whenever possible. When it was inconvenient to use the air cell, the stopping power of the foils was reduced by 5 %.

The energy of the second inelastic group could not be determined with such precision at each angle as in the case of the first group, because the low intensity made the exact position of the peak uncertain. The mean of the results at the different angles obtained with the foils gave a Q value of -1.89 MeV, while a single determination with the air cell gave -1.82 MeV. The best value for the excitation energy of this level would therefore seem to be 1.82 ± 0.04 MeV. In this calculation, it has been assumed that the level responsible for the group belongs to ^{24}Mg ; if, however, the isotope involved is either ^{25}Mg or ^{26}Mg , the excitation energy would be very slightly higher, at 1.83 or 1.84 MeV respectively. The fact that the amplitude of this peak is about one-tenth that of the first peak suggests that one of the less abundant isotopes may indeed be responsible, if the cross-section is assumed to be roughly the same for both levels. This is supported by the fact that no γ -rays involving a level of about 1.8 MeV excitation energy have been observed in the β -decay of ^{24}Na , though this argument is not conclusive because of the possibility that levels which may be excited by proton collision may be forbidden by selection rules in the process of β -disintegration. The presence of an inelastic group of protons at about this energy has recently been confirmed by Bush & Fulbright (1948), who give the excitation energy as 1.7 ± 0.3 MeV. These authors state that the intensity of this group is about the same as that of the first inelastic group, though their experiments were carried out at a much higher bombarding energy.

A brief investigation was made of the variation with energy of the ratio of the amplitudes of the inelastic and elastic groups. This was done by introducing an aluminium foil about 0.0004 in. thick into the path of the beam; the angle of the foil relative to the beam could be changed, so that the energy of the incident protons was continuously variable over a range of about 0.75 MeV. The ratio of the amplitude of the first inelastic group to that of the elastic group varied with incident energy as shown in table 1, the scattering angle being 90° . At lower energies the

TABLE 1. RATIO OF AMPLITUDES OF INELASTIC AND ELASTIC GROUPS

E_0 (MeV)	3.98	4.22	4.40	4.73
N_I/N_E	0.20 ± 0.03	0.11 ± 0.02	0.32 ± 0.04	0.68 ± 0.05

number of counts was very small because of the reduced intensity consequent on the small-angle scattering in the foil, which was some distance from the target. Nevertheless, the rise in the number of inelastic protons relative to the elastic protons at the lowest energy was outside the statistical error, and was reproducible. If the levels in the compound nucleus at the excitation energy corresponding to the incident proton energy are very closely spaced, we should expect the amplitude of the elastic group to remain more or less constant in the range 4.0 to 4.7 MeV, in which case the ratio of the two groups represents the excitation function of the

inelastic group. However, the binding energy of an additional proton to the ^{24}Mg nucleus is only 2.3 MeV, so that the levels of the compound nucleus involved will be only about 6 MeV above the ground state. At this energy the levels may be widely spaced, so that the process may be a resonant one, in which case the amplitude of the elastic group will not remain constant and the ratio of the two groups will not be a true representation of the excitation function of the inelastic group. This may well be the case in view of the fact that the ratio of the two groups shows a definite increase at the lowest energy, so that the above table cannot be relied upon to show more than the rough trend of the variation with energy of the inelastic group. The ratio of the amplitude of the second inelastic group to that of the elastic group was almost independent of energy in the range of 4.2 to 4.73 MeV.

The angular distribution of the elastically scattered protons was measured by operating the counter at very low bias, and inserting an absorbing foil sufficiently thick to cut out the inelastic group. Figure 6 shows the results expressed in the centre-of-mass co-ordinate system. The abscissa is the scattering angle, and the ordinate the relative number of protons N_ϕ at this angle multiplied by $\sin^4(\frac{1}{2}\phi)$ and

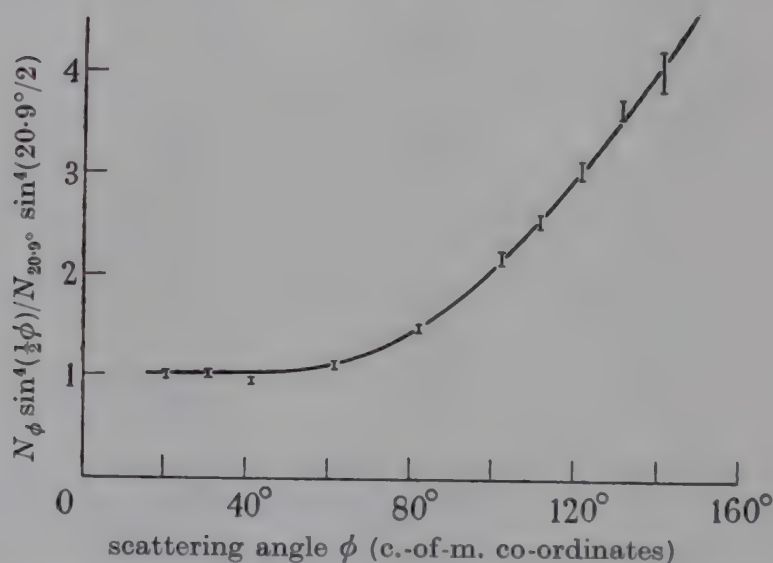


FIGURE 6. Elastic scattering from magnesium: ratio to intensity given by Rutherford's formula ($E_0 = 4.4$ MeV).

normalized to unity at 20.9° , which corresponds to 20° in the laboratory co-ordinate system. This curve therefore shows the ratio of the intensity at a scattering angle ϕ to the intensity that would exist if the scattering conformed to Rutherford's scattering formula, on the assumption that Rutherford's law holds at 20° . This assumption seems reasonable since the distance of closest approach of 4.7 MeV protons for a deviation of 20° by a pure Coulomb force is considerably greater than current estimates of the radius of the magnesium nucleus, and is supported by the fact that the ratio of the intensity to that given by Rutherford's formula remains constant up to about 40° . This indirect method of evaluating the results is not entirely satisfactory, but in the time available it was not possible to make a current integrator which gave reliable results, so that an absolute determination of the cross-section at each angle was not possible.

The ratio of the inelastic and elastic groups is given exactly by the ratio of the areas under the two peaks. This is not quite the same as the ratio of the heights of the two peaks, because of the difference in straggling, but the correction which has to be applied to the latter ratio to obtain the ratio of the numbers in the two groups stays sufficiently constant with angle for the relative heights of the peaks to be used as a measure of the angular distribution of the inelastic group. Figure 7 shows how the ratio of the inelastic to the elastic peak varies with angle. This method of measuring the ratio of the two groups broke down at small scattering angles be-

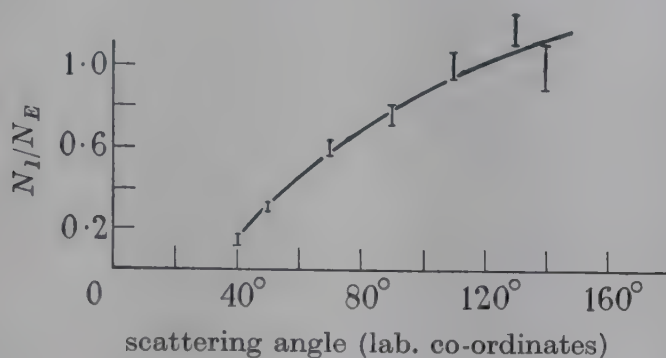


FIGURE 7. Ratio of amplitudes of first inelastic and elastic groups from magnesium ($E_0 = 4.73$ MeV).

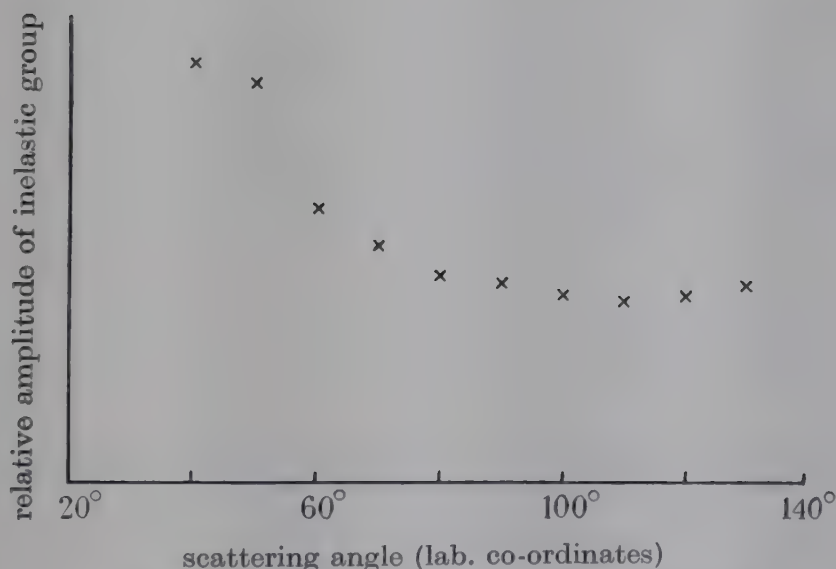


FIGURE 8. Angular distribution of first inelastic group from magnesium ($E_0 = 4.73$ MeV).

cause the background due to the elastic group became comparable with the inelastic group, making quantitative estimations of the latter impossible. This large background seemed to be due mainly to protons from the elastic group being scattered from the edges of the grid that was used to support the counter window, and losing energy in the process, so that they produced a larger ionization in the counter than they would otherwise have done. This background was much reduced when a counter with no grid was used, but no opportunity presented itself for repeating the experiment. An attempt to obtain the angular distribution of the inelastic group by multiplying the smoothed ratio of the peaks by the angular

dependence of the elastic group gave the points shown in figure 8. It seems unjustifiable to draw a curve through these points, because the rather complicated method of obtaining the information makes it almost impossible to estimate the error, but it is apparent that the inelastic group is more or less isotropic for angles greater than 80° , and shows a tendency to increase at forward angles. This behaviour resembles that observed by Heitler, May & Powell (1947) in the inelastic scattering of protons by neon, but contrasts with the results of Wrenshall (1943) for the inelastic scattering of 6.9 MeV protons by magnesium. The latter author found no inelastic protons whatsoever at 25° . It is unfortunate that no reliable results could be obtained at forward angles, as this region is particularly interesting from a theoretical standpoint.

An evaluation of the differential cross-section for inelastic scattering by comparison with the elastic group gave the value $6.5 \times 10^{-26} \text{ cm.}^2$ per unit solid angle at 90° . This refers to the first excited state at a bombarding energy of 4.73 MeV, and assumes it to be associated with ^{24}Mg . On the assumption that the inelastic group has a spherically symmetrical distribution, the total cross-section for excitation to the 1.38 MeV level is $8 \times 10^{-25} \text{ cm.}^2$. In view of the uncertainty in the angular distribution, this latter estimate is probably reliable only to within about 30 %.

(b) Aluminium

Aluminium was bombarded in the form of a thin foil of thickness 0.0005 in. (3.34 mg./cm.^2). Figure 9 shows the spectra of the scattered protons obtained at different angles, the energy of the incident protons being 4.57 MeV. The improved resolution shows clearly that the first inelastic group has a complex structure, and that the shape of the peak changes gradually between 90 and 130° . At 90° the peak is nearly resolved into two groups; below 90° the left-hand component seems to be

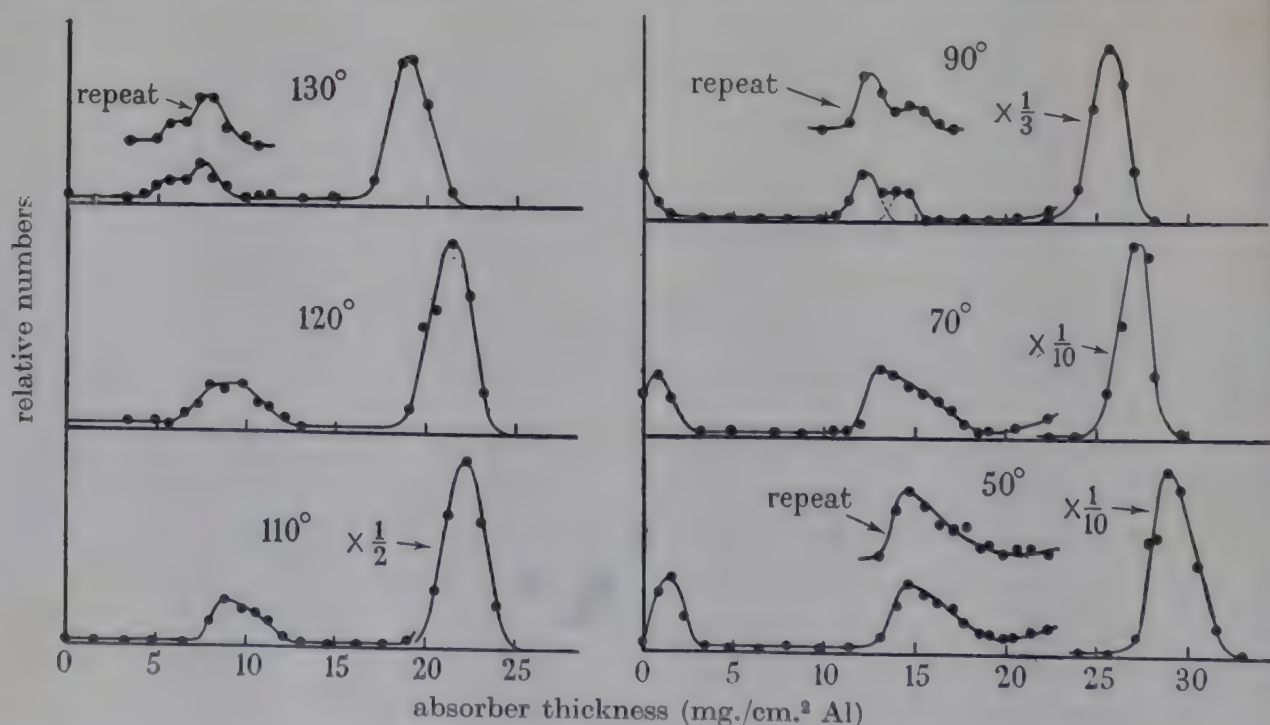


FIGURE 9. Spectra of protons scattered by aluminium ($E_0 = 4.58 \text{ MeV}$). The vertical scale is not necessarily the same at all angles.

consistently greater than the right, and the presence of the latter can be inferred only from the asymmetry of the peak, while above 90° the right-hand component appears to increase relative to the left-hand one, the two being approximately equal at 120° . Measurement of the Q values involved is difficult because of the complicated shape of the peak, but the position of the left-hand component seems fairly well established and is consistent with an excitation energy of 0.97 MeV at all angles. We may therefore regard the presence of an inelastic group corresponding to an excited state 0.97 ± 0.02 MeV above the ground state as certain. The energy of the right-hand component is much less definite. The shape of the peak at 90° is well explained by the presence of two levels at 0.97 and 0.80 MeV, but at other angles the shapes of the peaks are much too vague to allow of such a precise interpretation. All that can be said with certainty is that the level at 0.97 MeV is accompanied by at least one other level lying above 0.75 MeV. The second group of inelastically scattered protons was capable of penetrating the chamber and counter windows only at angles below 90° , and corresponds to an excitation energy of 2.15 ± 0.05 MeV. The above Q values were obtained with the aluminium absorbers, the stopping power being reduced by 5 % as mentioned above.

It is instructive to compare the results obtained in these experiments with those of other workers in the range below 2.5 MeV. Dicke & Marshall obtained levels at 0.87 and 2.03 MeV. Their value for the lower level probably represents the mean of the levels at 0.80 and 0.97 MeV quoted above, since their resolving power was inadequate to resolve the two groups. Their value for the other level is somewhat lower than 2.15 MeV, but since they obtain 1.32 for the 1.38 MeV level of magnesium, their results seem to be consistently about 5 % low. Davis & Hafner* made observations at 37 and 127° , and found levels at 0.84, 1.07, 1.79 and 2.28 MeV. On the assumption that their results are consistently about 5 % higher than those of the present experiments, the first two and the last levels agree closely with the results reported above. However, these workers never found the two lowest levels together at the same angle, nor, apparently, at the same bombarding energy; moreover, there is no trace of a group corresponding to a level at 1.79 MeV in the results shown in figure 9.

The angular distribution of the scattered protons was investigated as in the case of magnesium, the results being qualitatively similar. The elastic scattering gave a 'ratio-to-Rutherford' curve similar to that of figure 6, the Rutherford law being obeyed up to about 40° (at 4.2 MeV bombarding energy). The inelastic scattering was difficult to observe, because the complex nature of the first inelastic group made it impossible to estimate the amplitudes of the two components with any accuracy. The differential cross-section for excitation of the 0.97 MeV level by 4.38 MeV protons was about 10^{-26} cm.² per unit solid angle at 90° . On the assumption that the group has an approximately isotropic distribution, the total cross-section is about 1.2×10^{-25} cm.² This result is probably reliable to within 50 %. The ratio of the first inelastic to the elastic group increased only by about 10 % as the incident proton energy was raised from 4.38 to 4.73 MeV.

* The author is indebted to Dr E. M. Hafner for communicating his results before publication.

(c) *Copper*

Copper was bombarded in the form of a thin foil of thickness 2 mg./cm.², the incident proton energy being 4.73 MeV. No evidence of inelastic scattering was obtained. If an inelastic group corresponding to an excited state less than 2.5 MeV above the ground state was present, then its amplitude was less than 10⁻² times that of the elastic group.

(d) *Carbon and oxygen*

The scattering from a Cellophane target of thickness 0.001 in. was examined, the proton energy being again 4.73 MeV. Cellophane consists mainly of cellulose, and has the constitution (C₆H₁₀O₅)_n. The scattering from hydrogen is entirely elastic, and can be eliminated by observing at a scattering angle of 90°. At this angle there was no indication of any inelastic scattering, and it can be said with certainty that if there were present any groups of inelastically scattered protons corresponding to levels in either carbon or oxygen less than 2.5 MeV above the ground state, then the corresponding cross-section was at least 100 times smaller than that for elastic scattering. This confirms the work of May & Powell (1947), who reported no protons inelastically scattered by carbon at a proton energy of 4.2 MeV, and of Heitler *et al.* (1947), who found no inelastic groups from oxygen at the same energy.

At forward angles the group due to elastic scattering from hydrogen was able to reach the counter. This gave an independent check on the stopping power of the aluminium foils, since the energy of the scattered protons is calculable from the incident energy and the scattering angle. It was found that the stopping power of the foils was about 4% less than that calculated from the data of Livingston & Bethe, which agrees with the comparison between the foils and the air cell.

(e) *Beryllium*

A target was made by evaporating beryllium on to a copper foil, since it had already been ascertained that copper gave no inelastic scattering. The spectrum of the protons scattered through 40° is shown in figure 10. The groups are assigned to

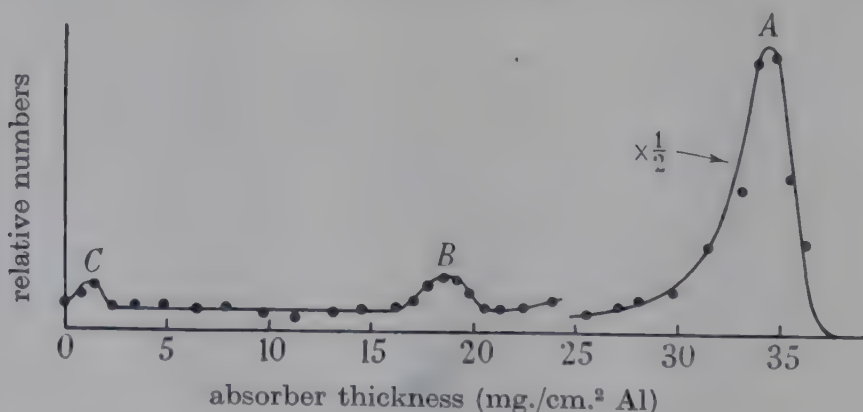


FIGURE 10. Spectrum of protons scattered through 40° by beryllium on copper foil ($E_0 = 4.57$ MeV).

the following reactions. Group *A* is clearly due to elastic scattering from copper and beryllium, the two components of which could be resolved at backward angles. Group *B* has the correct energy to be due to deuterons from the reaction ${}^9\text{Be}(p, d){}^8\text{Be}$

the Q of which has been measured as 0.547 MeV by Allison, Skaggs & Smith (1940). The group also showed a variation of range with angle consistent with this interpretation. Group C is more difficult to assign, since its range was sufficient to reach the counter only at forward angles, and the dependence of its energy on angle could not therefore be investigated. It could not be due to α -particles from the (p, α) reaction, since the Q of this process has been measured by Allison and his co-workers as 2.115 MeV, which is insufficient to account for the observed range. If the group consists of short-range deuterons from the (p, d) reaction, then the excited state in ^8Be involved would be 2.35 MeV above the ground state. According to Hornyak & Lauritsen (1948), the first excited state in ^8Be lies at 3 MeV, and this would have to be appreciably in error if the above explanation of group C were correct. The only remaining interpretation seems to be that the group consists of inelastically scattered protons; if this is correct, the excitation energy of the level involved is 2.39 ± 0.05 MeV. This interpretation agrees with the work of Davis & Hafner (1948*b*), who give the excitation energy as 2.41 MeV. Since the threshold for the (γ, n) reaction in ^9Be is 1.63 MeV, the level responsible for the inelastic scattering is unstable with regard to neutron emission.

DISCUSSION

(a) Spacing of the levels

It is convenient to collate here the excited states of nuclei which have been discovered through the inelastic scattering of protons. Table 2 contains the results of the present experiments together with those of other workers. The results of Dicke & Marshall have been increased by 5 % to make them agree where they overlap. In

TABLE 2. EXCITED STATES REVEALED BY INELASTIC SCATTERING

nucleus	excited states in MeV above ground state						
^9Be	2.39						
Mg	1.38, (1.82), (2.87),*	4.07,*	5.51,†	7.32,†	8.30†		
^{27}Al	0.80 (?), 0.97, 1.79,‡	2.15, 2.83,*	3.07,‡	3.20,‡	3.67*		
^{32}S	2.36,*	4.55*					
^{52}Cr	1.38,*	2.98,*	3.88*				

* Dicke & Marshall. † Fulbright & Bush. ‡ Davis & Hafner.

the case of magnesium, those levels enclosed in brackets are levels which, in view of their intensity, may belong to an isotope other than ^{24}Mg .

The information is too meagre to bear detailed analysis, but it is noticeable that the spacing of the levels is greater and decreases more slowly with excitation energy and with mass number than is predicted by most nuclear models. Before one can draw any definite conclusions about the spacings of the levels, it is necessary to consider the possibility that not all levels can be excited by proton bombardment. This is almost certainly true in view of the indifferent correlation between the energy levels obtained from inelastic scattering data and those found in β -disintegration. In the case of ^{24}Mg , the agreement is good. Wiedenbeck (1947) has established that the 1.38 and 2.76 MeV γ -rays emitted following the β -decay of ^{24}Na are in cascade,

which is compatible with levels at 1.38 and 4.14 MeV. (There is no sign of a γ -ray corresponding to a level at 1.82 MeV, but this may not be associated with ^{24}Mg .) For ^{27}Al , on the other hand, agreement is not so good. Itoh (1941) has reported the emission of γ -rays of 0.64, 0.84 and 1.01 MeV following the β -decay of ^{27}Mg . Bleuler & Zünti (1947), on somewhat obscure grounds, assign them to levels at 0.84 and 1.01 MeV, which agrees well with the first two levels found in inelastic scattering. However, the later work of Benes, Hedgran & Hole (1948) has established that the 0.84 and 1.01 MeV γ -rays are in cascade. These authors assign them to levels at 0.84 and 1.85 MeV, and their experiments seem the most reliable to date. If this be correct, we may identify the level at 0.84 MeV with that responsible for the high-energy component of the first inelastic group. The level at 1.85 MeV may be responsible for the group whose excitation energy is given by Davis & Hafner as 1.79 MeV. This group was observed at a bombarding energy of 5.5 MeV; there is no trace of it at 4.7 MeV in the experiments described here, and it is clear that, if the results of Benes and his co-workers be accepted, the levels obtained from β -decay do not always show up in inelastic scattering experiments, and vice versa. That the two sorts of experiment should reveal different excited states is hardly surprising, since the nuclear processes concerned are quite different and involve different selection rules. The density of levels may therefore be greater than that revealed by inelastic scattering (or, for that matter, by any other single experiment), but it is doubtful whether this is sufficient to explain the difference between the theoretical and observed level spacings.

(b) *Mechanism of the process*

The most obvious explanation of the inelastic scattering is that it is simply a (p, p) process in which the incident proton forms a compound nucleus and is re-emitted with less than its initial energy, leaving the residual nucleus in an excited state. However, Weisskopf (1938) and Guth (1945) have independently suggested that there may be a subsidiary process in which the nucleus is excited by the electric field of the approaching proton. Although both authors have given expressions for the cross-section of such a process, a comparison of their results with experiment is difficult because both formulae involve the electric moments of the corresponding γ -ray transitions, which at present are very uncertain. A comparison with the analogous process of nuclear excitation by electrons suggests that, for the conditions of the present experiment, the effect of the electric interaction is negligible compared with that of nuclear forces, though it may make some contribution to the cross-section at low bombarding energies.* This might be expected on general considerations, since for $Z = 12$ a 4.5 MeV proton can pass well over the top of the barrier. The cross-section for formation of the compound nucleus is then πR^2 , where R is the nuclear radius, and in the case of magnesium a rough explanation of the cross-section for inelastic scattering can be obtained on the assumption that the probabilities of the outgoing proton leaving the residual nucleus in the ground state and in an excited state are about equal.

* The author is indebted to Mr B. Touschek for discussion on this point.

If the process does involve formation of a compound nucleus, it might be expected that the existence of inelastic scattering would depend on the presence or otherwise of alternative modes of disintegration. In particular, we might expect that for medium values of Z the potential barrier would favour the emission of a neutron rather than a proton. It is of interest, therefore, to correlate the properties with regard to inelastic scattering of those nuclei from Mg onwards with the corresponding (p, n) threshold. This is shown in table 3. It is almost certain that all

TABLE 3. CORRELATION BETWEEN PROTON SCATTERING AND (p, n) THRESHOLD

nucleus	inelastic scattering?	(p, n) threshold (MeV)
^{24}Mg	yes	13
^{27}Al	yes	6
^{32}S	yes*	no (p, n) reaction known
^{52}Cr	yes*	6.4
$^{58, 60}\text{Ni}$	no*	4
$^{63, 65}\text{Cu}$	no	4.1, 2.2
$^{64, 66, 68}\text{Zn}$	no*	4.1, 4.2, 3.7

* Data obtained by Dicke & Marshall at 6.9 MeV.

nuclei with $A > 24$ must have at least one level under 4 MeV, so the absence of inelastic scattering in Ni, Cu and Zn is ascribable to the proton energy being above the (p, n) threshold. It is true, of course, that the nuclei which show no inelastic scattering are also those with high values of Z , but 6.9 MeV protons should be able to penetrate the barrier even of zinc, and only the variation of the (p, n) threshold would seem to explain why chromium gives inelastic scattering but copper does not.

It might be expected that the excitation function of the inelastic group would be dependent on the mechanism of the process. The data given previously on the variation with energy of the ratio of the inelastic to the elastic group can only be taken to show the general trend of the absolute cross-section, but it is evident that the group corresponding to the 1.38 MeV level in magnesium increases rapidly with proton energy in the region 4.2 to 4.7 MeV, whereas the group corresponding to the 0.97 MeV level in aluminium is almost independent of bombarding energy in this range. If the process is one involving formation of a compound nucleus, this behaviour can be understood in terms of penetration of the potential barrier by the outgoing proton, since, in the case of aluminium, the emitted proton has almost enough energy to pass over the top of the barrier at the lowest bombarding energy used. On the other hand, this simple picture does not explain why the group corresponding to the 1.82 MeV level in magnesium does not also show a sharp rise with increasing proton energy.

(c) Angular distribution

A detailed analysis of the angular distribution of the elastically scattered protons is difficult, and requires information at several energies. It might be expected that in the case of magnesium and aluminium the scattering would obey Rutherford's law up to about 40° , since for magnesium the distance of closest approach of 4.5 MeV protons scattered through 40° by pure Coulomb forces is 7.5×10^{-13} cm., which is just about equal to the nuclear radius plus the range of the nuclear forces. The

general shape of the 'ratio-to-Rutherford' curve for magnesium, shown in figure 6, is similar to that of the theoretical curves given by Heitler *et al.* for carbon, nitrogen and oxygen. These curves are calculated from the formula given by Mott & Massey (1931) on the basis of the one-body model, and take into account the *S*-wave anomaly only. However, a detailed analysis along these lines seems hardly worth while, since Bethe (1937) has pointed out that the one-body theory gives incorrect results if the spins of nucleus or particle are different from zero, and if inelastic scattering has a probability comparable with that of the elastic process.

The angular distribution of the inelastically scattered protons is also complicated. Since the process may be regarded as a (p, p) reaction, the problem is similar to that of the angular distribution of disintegration products, with the added complication that not only the intermediate states but also the final states are unknown, since they depend on the spin of the excited state involved. Fortunately, we can make use of a theorem enunciated by Eisner & Sachs (1947), which states that 'in a nuclear reaction produced with an unpolarized beam of given orbital angular momentum incident on an unpolarized target, the angular distribution of the outgoing intensity cannot be more complicated than that of the incoming intensity'. This means that an incident *S*-wave can only give rise to an isotropic distribution of disintegration particles. From this it is plausible to assume that the first inelastic group from magnesium, which has an approximately isotropic distribution, arises mainly from incident protons having zero angular momentum. The case of the first inelastic group from aluminium is more complicated, since the amplitudes of the two components show a marked angular dependence. The angular distribution depends in a complex manner on many factors, among them the spin of the residual nucleus. This suggests that the two components belong to closely situated levels having different spins; these levels may form a 'doublet' in the spectroscopic sense.

In conclusion, it is interesting to consider the effect on the angular distribution of excitation by electric forces. It might be expected that a long-range force of this sort would give rise to increased scattering at small angles. It is perhaps easier to see this by considering the converse; if the scattering increases rapidly at small angles, then the outgoing wave function must contain spherical harmonics of high order, and the incident protons responsible must have very high angular momentum, because of the theorem enunciated above. This high angular momentum would be possible only if the forces involved were of long range. The experimental evidence on this point is not very consistent. Heitler *et al.* found that the angular distribution of the protons inelastically scattered by neon at an incident energy of 4.2 MeV showed a sharp rise at angles less than 40° . On the other hand, Wrenshall, who studied the scattering of 6.9 MeV protons by magnesium, found a continuous decrease in the number of inelastically scattered protons as the scattering angle decreased from 90° , and states categorically that none was found at 20° . The results of the present experiments are somewhat inconclusive, since observations could not be made below 40° , but there was certainly no decrease in intensity of the elastic group at forward angles, and at 40° it showed a tendency to rise, which resembles the behaviour observed in the case of neon. This point seems of sufficient interest to warrant further investigation with an improved technique.

The author would like to thank Mr L. Coates and Dr A. E. Kempton for assistance in the experiments, and to express his indebtedness to the latter for much helpful discussion throughout the work. He would also like to acknowledge the help given by members of the Cambridge cyclotron team, without which the experiments would have been impossible, and to thank Professor P. I. Dee and Mr B. Touschek for criticizing the manuscript. The work was carried out while the author was in receipt of a maintenance grant from the Department of Scientific and Industrial Research.

REFERENCES

- Allison, S. K., Skaggs, L. S. & Smith, N. M. 1940 *Phys. Rev.* **57**, 550.
 Benes, J., Hedgran, A. & Hole, N. 1948 *Ark. Mat. Astr. Fys.* **35A**, paper 12.
 Bethe, H. A. 1937 *Rev. Mod. Phys.* **9**, 69.
 Bleuler, E. & Zünti, W. 1947 *Helv. Phys. Acta*, **20**, 195.
 Bush, R. R. & Fulbright, H. W. 1948 *Phys. Rev.* **74**, 1206.
 Creutz, E. C. & Wilson, R. R. 1946 *Rev. Sci. Instrum.* **17**, 385.
 Davis, K. E. & Hafner, E. M. 1948*a* *Phys. Rev.* **73**, 1242.
 Davis, K. E. & Hafner, E. M. 1948*b* *Phys. Rev.* **73**, 1473.
 Dicke, R. H. & Marshall, J. 1943 *Phys. Rev.* **63**, 86.
 Eisner, E. & Sachs, R. G. 1947 *Phys. Rev.* **72**, 680.
 Fulbright, H. W. & Bush, R. R. 1948 *Phys. Rev.* **74**, 1323.
 Guth, E. 1945 *Phys. Rev.* **68**, 280.
 Heitler, H., May, A. N. & Powell, C. F. 1947 *Proc. Roy. Soc. A*, **190**, 180.
 Hornyak, W. F. & Lauritsen, T. 1948 *Rev. Mod. Phys.* **20**, 191.
 Itoh, J. 1941 *Proc. Phys. Math. Soc. Japan*, **23**, 605.
 Livingston, M. S. & Bethe, H. A. 1937 *Rev. Mod. Phys.* **9**, 245.
 May, A. N. & Powell, C. F. 1947 *Proc. Roy. Soc. A*, **190**, 170.
 Mott, N. F. & Massey, H. S. W. 1931 *Theory of atomic collisions*, p. 275. Oxford University Press.
 Powell, C. F., May, A. N., Chadwick, J. & Pickavance, T. G. 1940 *Nature*, **145**, 893.
 Rhoderick, E. H. 1948 Ph.D. Thesis, Cambridge University.
 Rhoderick, E. H. 1949 *Nature*, **163**, 848.
 Siegbahn, K. 1946 *Phys. Rev.* **70**, 127.
 Weisskopf, V. F. 1938 *Phys. Rev.* **53**, 1018.
 Wiedenbeck, M. L. 1947 *Phys. Rev.* **72**, 429.
 Wilkins, T. R. 1941 *Phys. Rev.* **60**, 365.
 Wrenshall, G. 1943 *Phys. Rev.* **63**, 56.

Paramagnetic resonance in copper sulphate

BY D. M. S. BAGGULEY AND J. H. E. GRIFFITHS, *Clarendon Laboratory,
University of Oxford*

(Communicated by Lord Cherwell, F.R.S.—Received 15 October 1949—

Revised 12 December 1949)

Paramagnetic resonance absorption in single crystals of copper sulphate pentahydrate has been investigated at wave-lengths of 3.04, 1.23 and 0.85 cm. Only one absorption line is found at wave-lengths of 3.04 and 1.23 cm., and the position agrees with that calculated from susceptibility data. Two lines would be expected theoretically for many orientations of the crystal in the magnetic field, corresponding to the two inequivalent Cu^{++} ions in unit cell. At 0.85 cm. wave-length two lines are found under the appropriate conditions but not at longer wave-lengths. This is interpreted in terms of exchange interaction between dissimilar ions, and the exchange frequency is estimated at 0.15 cm.^{-1} . Other effects of exchange interaction in this salt are discussed.

1. INTRODUCTION

The resonant absorption of microwave radiation in paramagnetic salts placed in a magnetic field was first reported by Zavoisky (1945, 1946), who observed the absorption in powdered samples of MnSO_4 , etc. Further work has been done on powders, but it is by the use of single crystals that it is possible to obtain detailed information about the lowest energy levels of the paramagnetic ion by this method. The chrome alums, in particular, have been investigated in some detail (Bagguley & Griffiths 1947; Bleaney 1949; Halliday & Wheatley 1948; Whitmer, Weidner, Hsiang & Weiss 1948). In general, there are two points of interest which may be investigated, the separations of the lowest energy levels, which may be correlated with the crystalline field theory for the particular ion, and the width of the absorption line. This latter may give information about the interactions of the ions with each other and with the lattice. A general survey of the results obtained with salts of the iron group of transition elements is given by Bagguley, Bleaney, Griffiths, Penrose & Plumpton (1948).

In this paper are described measurements at room temperature on single crystals of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ at wave-lengths of 3.04, 1.23 and 0.85 cm. A preliminary account of this work has been given by Bagguley & Griffiths (1948), and some measurements at 3.25 cm. have been reported by Arnold & Kip (1949) and also by Wheatley & Halliday (1949) which agree with the results described here. The results are compared with the crystalline field theory worked out by Polder (1942) for this salt, and with susceptibility measurements. There is general agreement between the results of the absorption measurements and those of the susceptibility measurements, but the absorption measurements show two features of considerable interest. The first is that the line width for certain orientations of the magnetic field is much smaller than would be expected from the normal magnetic interactions of the ions. Secondly, two lines would be expected theoretically for many orientations of the crystal in

the magnetic field, corresponding to the two inequivalent Cu^{++} ions in unit cell. These are observed at 0.85 cm. wave-length but not at longer wave-lengths.

Gorter & Van Vleck (1947) have suggested that the effect of exchange interaction is to narrow the absorption line, and it appears likely that this effect is observed in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$. The appearance of only one line can also be interpreted in terms of exchange interaction, in this case between dissimilar ions. This is discussed further in § 6.

2. EXPERIMENTAL ARRANGEMENT

The apparatus used in these experiments will be described in another paper, which is in course of preparation, and it is sufficient to give only a brief description here, together with such details as are particularly relevant to these investigations.

The oscillators used were of the low-power (12 to 100 mW output) reflexion klystron type; at 3 cm. a CV 87, at 1.23 cm. a VX 302 and at 0.85 cm. a special valve of the same type as the VX 302. The voltage supplies to the oscillator were carefully stabilized, since it was found that frequency or amplitude fluctuations in the valve were the main limitations on the accuracy and sensitivity of the method.

Power from the oscillator was fed through a wave-guide 2 m. long to a cylindrical resonator with at least 7 db. of attenuation inserted between the oscillator and the resonator. This length of guide was found sufficient to make any effect of the magnetic field on the valve negligible. The feed into the resonator was a small hole in the side about a quarter of a cavity wave-length from the end and another hole of the same size opposite the first, fed power out of the resonator into a short length of guide terminated by a silicon crystal detector unit. The resonators used had diameters 22 mm. at 3.04 cm. and 1.23 cm. wave-length and 8.5 mm. at 0.85 cm. wave-length. The resonators were excited in the H_{11} mode at 3.04 and 0.85 cm. wave-length and in the H_{01} mode at 1.23 cm. The top of the resonator was closed by a non-contact tuning plunger supported from a micrometer screw and the bottom by another non-contact plunger which could be rotated and on which the crystal was placed. The resonator was placed between the poles of an electromagnet so that the field was at right angles to the h.f. magnetic field in the resonator.

Measurements were made at constant frequency by tuning the resonator to resonance with the copper sulphate crystal in position and taking readings of the detector current at different magnetic field strengths, the resonator being tuned before each reading. Since the detector current did not exceed $2 \mu\text{A}$ the rectification law can be assumed to be a square law, and in this case the increased absorption in the sample due to the paramagnetic resonance is proportional to

$$\left(\frac{Q_0}{Q_H} - 1\right) = \left(\sqrt{\left(\frac{\theta_0}{\theta}\right)} - 1\right),$$

where Q_0 and θ_0 are the magnification factor of the resonator and the galvanometer reading, respectively, at zero magnetic field and Q_H and θ are these quantities at a magnetic field H . In all results given in this paper the absorption is expressed in arbitrary units, and no attempt has been made to measure the absolute value.

The magnetic field was measured by a search coil calibrated by the N.P.L. and a fluxmeter which was checked against a ballistic galvanometer using a substandard

ammeter and mutual inductance. The accuracy of the measurement of magnetic field was from $\frac{1}{2}$ to 1 %. For the very narrow lines this accuracy was not sufficient to obtain the shape of the line and another procedure was used. The magnetic field was adjusted to be at the centre of the absorption line at the start of each measurement. A small change in the magnetic field was then made and this change was measured by a ballistic galvanometer connected to a coil permanently in the field. This procedure was repeated for different changes in the field until the line had been delineated. An example of a curve obtained in this way is shown in figure 6.

3. CRYSTALLOGRAPHIC AND SUSCEPTIBILITY DATA

The analysis of the crystal structure of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ given by Beevers & Lipson (1934) shows that there are two inequivalent Cu^{++} ions in a unit cell, with parameters (000), and $(\frac{1}{2}\frac{1}{2}0)$. Each ion is surrounded by a nearly octahedral arrangement consisting of four water molecules and two oxygen atoms; the angle between the planes of water molecules of the two ions is 82° . Using these data Polder (1942) has calculated the effect of the crystalline electric field on the Cu^{++} ions. He takes an electric field of cubic symmetry with a strong component of tetragonal symmetry, the tetragonal axis being normal to the plane of the water molecules. The Cu^{++} ion is in a D state which is split by the cubic field into a triplet and a doublet, the doublet lying lowest. The tetragonal field removes the orbital degeneracy, giving a separation between the two components of the doublet of the order of $12,000 \text{ cm}^{-1}$. Thus the lowest orbital level is well removed from the next level.

The spin degeneracy is removed by a magnetic field, giving a doublet whose separation is $g\beta H$, where H is the magnetic field, β the Bohr magneton and g depends on the orientation of the tetragonal axis with respect to the magnetic field. For a completely free spin g would be 2, and the effect of the spin-orbit coupling is to increase g to a value of 2.47 when H is along the tetragonal axis and 2.06 when H is at right angles to this axis. When account is taken of the two Cu^{++} ions, this gives good agreement with the results of the susceptibility measurements of Krishnan & Mookherji (1936, 1938). They find that the crystal is magnetically uniaxial, with the principal axis parallel to the intersection of the planes of the water molecules of the two ions, and that in this direction the susceptibility is a minimum. The values of the susceptibilities given by these authors parallel and perpendicular to the magnetic axis are $\chi_{\parallel} = \frac{0.399}{(T-2.0)}$ and $\chi_{\perp} = \frac{0.486}{(T+1.8)}$, giving effective g values of 2.07 and 2.26.

Paramagnetic resonance occurs when the radiation energy $h\nu$ is equal to the energy difference between the spin levels, i.e. in this case when $h\nu = g\beta H$. Thus since g is dependent on the orientation of the magnetic field, one would expect the position of the absorption line to be similarly dependent and also for there to be two lines corresponding to the two Cu^{++} ions except when the g values of the two ions are equal or are not sufficiently different for the two lines to be resolved.

Another theoretical treatment of the susceptibility of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ was given by Jordahl (1934) in an earlier paper and with less accurate data available. He considered only one ion per unit cell and obtained a very much smaller splitting (about

200 cm.^{-1}) for the basic orbital doublet, so that both levels would be occupied at room temperature. Since there are a number of discrepancies between the results of this treatment and both the susceptibility and the absorption data it will not be considered further in this paper.

4. RESULTS

Most of the measurements were made at three different orientations of the crystal in the magnetic field. The first, denoted by L_1 , is with the field parallel to the principal axis of magnetic susceptibility. This is at right angles to the tetragonal axes of both ions and makes angles of (154° , 64° , 51°) with the axes of unit cell. The second, L_2 (78° , 130° , 52°), is along the tetragonal axis of the ion at (000) and the third L_3 is the internal bisector of the two tetragonal axes L_2 and L_4 (70° , 41° , 69°).

Plane sections of from 1 to 2 mm. thick were ground from a crystal, the plane of section I containing L_1 and L_2 and of section II containing L_2 , L_3 and L_4 . The results for the field parallel to the directions L_1 , L_2 and L_3 are given in table 1. Here g is obtained from the relation $g\beta H = h\nu$, and the values of g (susceptibility) given for comparison are taken from Krishnan & Mookherji (1936), except the values for L_2 at 0.85 cm. which are referred to later. $\Delta H_{\frac{1}{2}}$ is the difference in magnetic field between the position of maximum absorption and that at which the absorption is half its maximum value.

The accuracy of the g values is about $\pm \frac{1}{2}\%$ except for the very wide lines where it is somewhat less.

TABLE 1

direction of H	L_1 (154° , 64° , 51°)			L_2 (78° , 130° , 52°)				L_3 (70° , 41° , 69°)		
wave-length (cm.)	3.04	1.23	0.85	3.04	1.23	0.85		3.04	1.23	0.85
H at resonance (kG)	3.36	8.36	12.1	3.11	7.64	10.6	11.9	3.10	7.70	11.2
g	2.09	2.08	2.08	2.26	2.27	2.37	2.12	2.27	2.26	2.25
g (susceptibility)	2.06	2.06	2.06	2.27	2.27	2.47	2.06	2.27	2.27	2.27
width $\Delta H_{\frac{1}{2}}$ (gauss)	28	24	25	115	640	$\simeq 450$		58	50	—

5. DISCUSSION OF RESULTS

(a) g values

The agreement between the g values obtained by this method and those obtained from susceptibility measurements is within the experimental errors. In the case of L_2 at 0.85 cm. the mean of the two g values is 2.25, again in agreement with the susceptibility value. On the theory outlined in § 3, two lines should be observed in this direction and the g values should be 2.06 for the ion whose tetragonal axis is at right angles to the magnetic field, since this is the same as when H is along L_1 , and 2.47 for the other ion to give a mean value of 2.27. These are the values given in table 1.

At 3.04 cm., only one line is observed with H along L_2 . The experimental curve for this case is shown in figure 1 and the expected positions of the two lines are marked at 2.86 and 3.42 kG. From this it is clear that it is not because of lack of

resolution that the two lines are not observed. Figure 2 shows the two lines resolved at 0.85 cm. The greater scatter of the experimental points in this case is due to the fact that the valve used at this wave-length was not as steady as the other valves.

A number of measurements were made at 3.04 cm. wave-length at various directions in the plane $L_2L_3L_4$ which contains the tetragonal axes of the two ions.

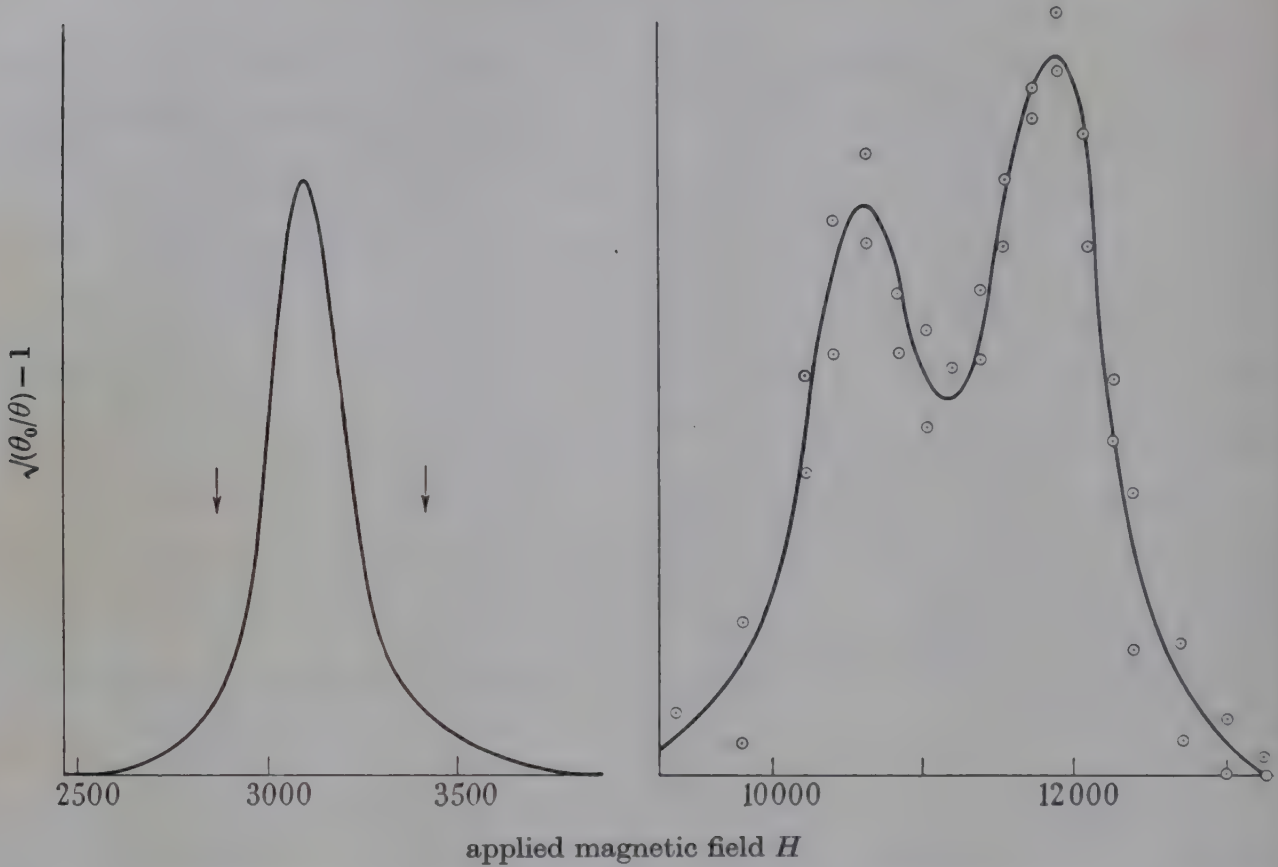


FIGURE 1. Absorption curves for H along L_2 at $\lambda = 3.04$ cm. FIGURE 2. Absorption curve for H along L_3 at $\lambda = 0.85$ cm.

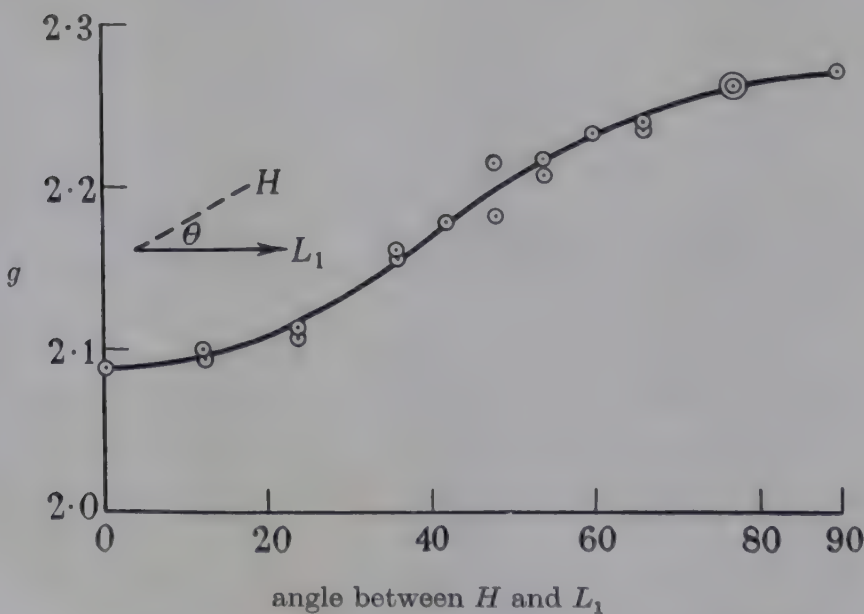


FIGURE 3. Variation of g with orientation for H in the plane L_1L_2 at $\lambda = 3.04$ cm.

If the angle between these axes is $(90 - \alpha)$ and θ is the angle between H and one of the axes, the mean g value measured will be given by

$$g^2 = \frac{1}{2}(g_{\parallel}^2 + g_{\perp}^2) + \frac{1}{4}(g_{\parallel}^2 - g_{\perp}^2) \sin 2\alpha \sin 2\theta,$$

where $g_{\parallel}$ and $g_{\perp}$ are the g values of an ion when H is respectively parallel and perpendicular to the tetragonal axis. No variation of the g value greater than $\frac{1}{2}\%$ (the experimental error) was found, which shows that α is less than 2° . This agrees with the susceptibility measurements of Krishnan and shows that the tetragonal axes of the electric field are more nearly at right angles to each other than is suggested by the crystallographic data.

Measurements were also made in the L_1L_2 plane at 3.04 cm. wave-length, and the results are shown in figure 3 in which the points are the experimental results and the full curve the theoretical values given by the expression $g = g_{\parallel} + (g_{\perp} - g_{\parallel}) \sin^2 \theta$, where it is assumed that $(g_{\perp} - g_{\parallel})$ is small. Again the agreement is within the experimental error.

(b) Width of lines

Great variation of the width of the line was found both with orientation and with wave-length. For H parallel to L_1 , the width is very small and appears to be independent of wave-length over the range investigated. Similarly, when H is parallel to L_3 there is no very marked change of width with wave-length. For both these directions the magnetic field is equally inclined to the tetragonal axes of the two Cu^{++} ions and therefore the effective g value is the same for each ion.

For L_2 where the g values of the two ions should be different the width increases rapidly with decreasing wave-length until the two lines are separated. Even at

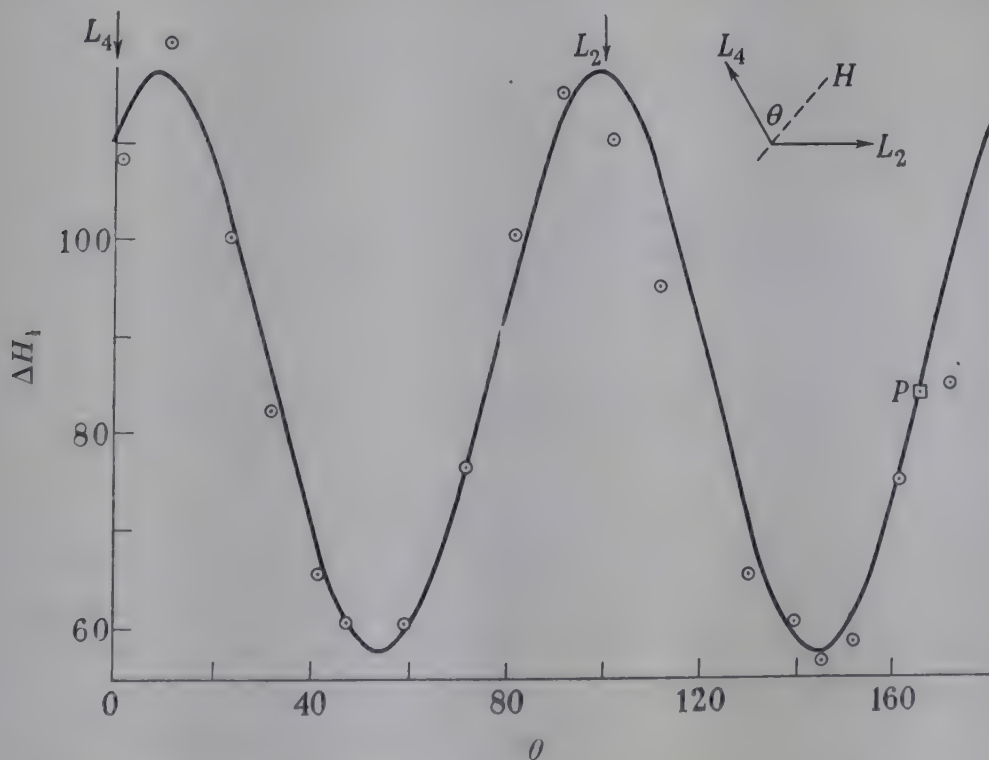


FIGURE 4. Variation of width with orientation for H in the plane L_2L_4 at $\lambda = 3.04$ cm.

1.23 cm. wave-length there is some indication of the two lines, since the top of the absorption curve is very flat.

The variation of width with orientation in the plane L_2L_4 is shown in figure 4 and in the plane L_1L_2 in figure 5. The measurements of figure 4 were made at 3.04 cm. and those of figure 5 at 1.23 cm. It was suspected that small errors in the grinding and marking of the crystal section used for the measurements of figure 4 might lead to an appreciable error in the final orientation of the crystal. A check was therefore made with an uncut crystal oriented with H at 13.5° to L_4 (point P of figure 4). The choice of this orientation was made because the width is sensitive to change of orientation in this region, and, secondly, it is easy to set an uncut crystal accurately in this direction, since it is the intersection of the plane L_2L_4 with the plane of the crystal face μ and lies at 79° from the c axis of the crystal.

This measurement showed that there was an error of 8° in the original set of measurements and the curve of figure 4 has been adjusted accordingly. This error was found to be mostly in the marking of the line L_2 in the plane L_2L_4 and not in the cutting of the plane.

The shape of the resonance line with H along L_1 measured at 1.23 cm. is shown in figure 6. Comparison with a Gaussian curve shows that the wings of the absorption line have a considerably greater intensity than in the case of the Gaussian curve.

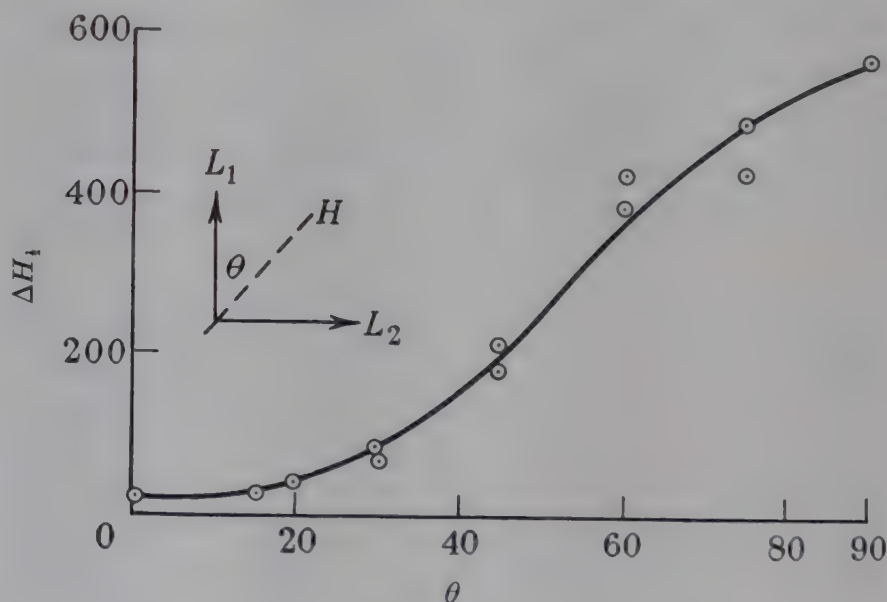


FIGURE 5. Variation of width with orientation for H in the plane L_1L_2 at $\lambda = 1.23$ cm.

6. EFFECT OF EXCHANGE

The width of the absorption line should be due either to spin-lattice interaction or spin-spin interaction or a combination of both effects (Bagguley *et al.* 1948). In this case the spin-lattice relaxation time is difficult to estimate. Paramagnetic dispersion experiments (Gorter 1947) give a relaxation time of the order of 3×10^{-8} sec. at liquid-air temperature, which would give a line breadth of about 10 G. If this relaxation time decreases rapidly with increase of temperature, as in some substances, it would lead to a very considerable broadening of the line at room tem-

peratures. However, since the narrowest line measured here is only 25 G and preliminary experiments show no great change of width with temperature, it does not seem likely that spin-lattice interaction is affecting the width to any marked extent.

The spin-spin interaction has been investigated theoretically by Van Vleck (1948) and by Pryce & Stevens (1950) and, for the case of nuclear spins, by Bloembergen, Purcell & Pound (1948). They considered the effect of magnetic interaction between the ions, which results in a broadening of the absorption line. This effect can be represented as a fluctuating internal magnetic field whose root-mean-square value in the case of copper sulphate is calculated to be about 240 G. The width of the absorption line should be of this order, whereas in fact very much smaller widths are measured, down to 25 G.

There are therefore two main features of these results which are not in agreement with existing theory, namely, the occurrence of narrow lines and the failure to observe two lines at the longer wave-lengths. Gorter & Van Vleck (1947) have suggested that the narrowing of the line may be due to the effect of exchange interaction between the Cu^{++} ions. A rapid exchange of electrons between the ions will have the effect of smoothing the fluctuations in the magnetic field, an effect which is similar to the narrowing of the nuclear resonance line in a liquid due to the motion of the molecules (Bloembergen *et al.* 1948). This applies to similar ions, i.e. ions with the same energy in the magnetic field, but for dissimilar ions another effect will be introduced. In this case the precessional frequencies of the two ions are different, and if there is exchange, the type of absorption curve obtained will depend on whether the exchange frequency is greater or smaller than the difference in precessional frequencies. When it is greater, a mean value of the precessional frequencies should be measured, whereas when it is smaller, each ion should be observed separately and two lines should appear.

This qualitative description seems to cover the main facts. Thus with H along L_1 , all ions are similar and a narrow line results. With H along L_3 the precessional frequencies of the two types of ions are the same, but the precessional axes are slightly different, which may account for the rather greater width. The greatest difference between the ions is when H is along L_2 and this gives the greatest width. A very rough estimate of the exchange frequency can be made if it is assumed that the two lines are just not resolved when the exchange frequency is equal to the difference in the precessional frequencies. The measurement at 1.23 cm. wave-length shows some indication of the two lines and this may be taken as the appropriate wave-length. At this wave-length the maximum absorption with H along L_2 occurs at 7640 G, and in this field the difference in the resonance frequencies of the two ions is 0.15 cm.^{-1} . This is therefore the order of the exchange frequency.

A quantitative theory of the broadening due to exchange of dissimilar ions has been given by Pryce (1948), who gives an expression for the broadening due to this cause in terms of the g values and the orientation of the magnetic field. This expression contains two undetermined constants A and B which depend on the frequency but are independent of angle and temperature. So far it has not been possible to check this expression in any detail, and in fact this is somewhat difficult

to do, as it only applies at wave-lengths long compared with the resolving wave-length and then the variation in width is small. However, it does give the right form of variation, and with the empirical values $A = -0.91$, $B = 0.74$ gives a good fit with the experimental points of figure 4, although a $\cos^2 \theta$ curve gives an equally good fit. It is interesting to note that the variation of width with orientation in this plane ($L_2 L_3$) approximates to a $\cos^2 \theta$ form even at 1.23 cm. wave-length where the lines are very nearly resolved.

The behaviour of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ may be contrasted with that of the Tutton salts such as $\text{CuSO}_4 \cdot \text{K}_2\text{SO}_4 \cdot 6\text{H}_2\text{O}$ (Bleaney, Penrose & Plumpton 1949). In these salts the exchange frequency seems to be small, since the two lines are observed in the appropriate direction at 3 cm. wave-length giving g values in close agreement with theory.

7. DISCUSSION

It has been shown above that the g values obtained by experiments on paramagnetic resonance in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ are in agreement with the susceptibility measurements of Krishnan & Mookherji (1936), and also that their interpretation in terms of two inequivalent ions per unit cell is consistent with these results if it is assumed that there is an exchange interaction between these ions. Van Vleck (1948) has considered the case of exchange between similar ions. The quantities he calculates are (in his notation) $\langle \Delta\nu^2 \rangle_{\text{av.}}$ and $\langle \Delta\nu^4 \rangle_{\text{av.}}$ which can be defined by the relation

$$\langle \Delta\nu^n \rangle_{\text{av.}} = \frac{\int_{-\infty}^{\infty} A(\nu) (\nu - \nu_0)^n d\nu}{\int_{-\infty}^{\infty} A(\nu) d\nu},$$

where $A(\nu)$ is the intensity of absorption at frequency ν and ν_0 is the frequency of maximum absorption. He shows that $\langle \Delta\nu^2 \rangle_{\text{av.}}$ is not affected by exchange interaction, whereas there is a positive contribution to $\langle \Delta\nu^4 \rangle_{\text{av.}}$, which means that the half-width of the absorption line is decreased by exchange. For dissimilar ions, however, exchange does contribute to $\langle \Delta\nu^2 \rangle_{\text{av.}}$, and the half-width of the line is now increased. The effect of exchange on the width of the absorption lines is therefore twofold: a broadening when the ions are dissimilar and a narrowing when the ions are similar. The case of dissimilar ions has been considered above, and it was shown that an estimate of 0.15 cm.^{-1} can be made for the exchange frequency in that case.

Unfortunately, it has not been possible to estimate experimentally $\langle \Delta\nu^2 \rangle_{\text{av.}}$ and $\langle \Delta\nu^4 \rangle_{\text{av.}}$ for those orientations of the crystal in which all the ions are similar. The observations at present suggest a line shape of the form

$$A = 1/(1+x^2),$$

where x is the distance from resonance measured in units of the half width, so that the contribution from the wings of the absorption line is sufficient to make $\langle \Delta\nu^2 \rangle_{\text{av.}}$ and $\langle \Delta\nu^4 \rangle_{\text{av.}}$ infinite. It is necessary, therefore, for more sensitive experiments to be carried out in the wings of the line to determine the true values of these quantities. However, as stated above, this exchange narrowing is analogous to the narrowing

of the nuclear resonance line in liquids due to the motion of the magnetic dipoles, which has been treated by Bloembergen *et al.* (1948). Using this analogy, Van Vleck suggests that the half-width $\Delta\nu_{\frac{1}{2}}$ of the line should be of the order of $\tau_c \langle \Delta\nu^2 \rangle_{av.}$, where τ_c is the relaxation time for the process. The calculated value of $[\langle \Delta\nu^2 \rangle_{av.}]^{\frac{1}{2}}$ for copper sulphate taken from Van Vleck's paper is 0.02 cm.^{-1} , and the measured value of the width of the narrowest absorption line expressed as $\Delta\nu_{\frac{1}{2}}$ is $5 \cdot 10^{-3} \text{ cm.}^{-1}$ which gives a value for $1/\tau_c$ of 0.08 cm.^{-1} . If this is interpreted as the exchange frequency it is seen to be rather smaller than the value of 0.15 cm.^{-1} obtained above, although in view of the very rough approximation made, the discrepancy is not significant.

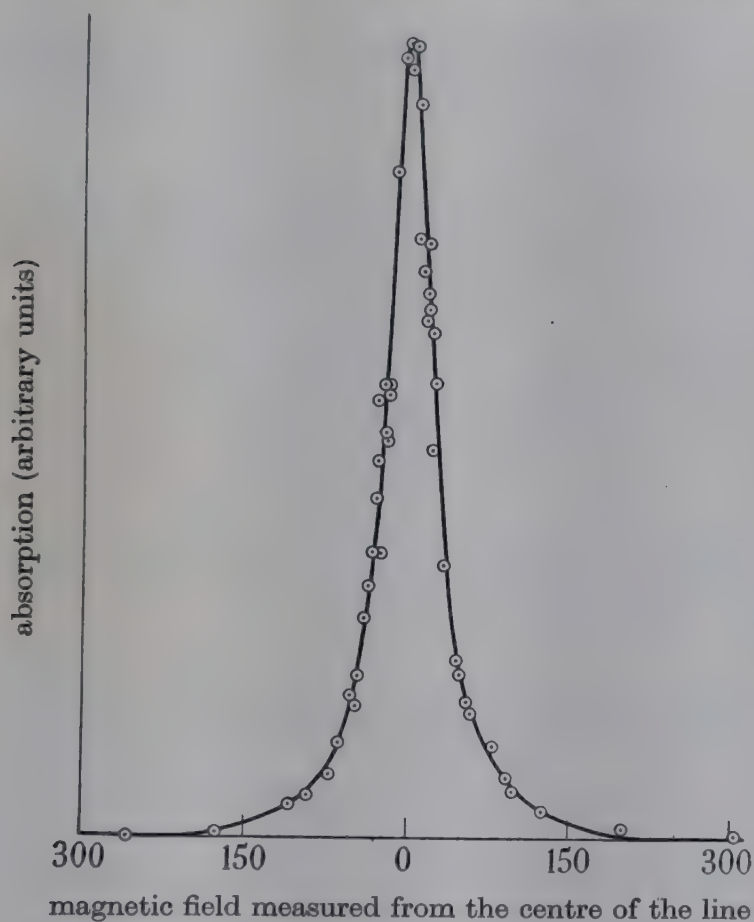


FIGURE 6. Absorption curve for H along L_1 at $\lambda = 1.23 \text{ cm.}$

Exchange interaction should also have an effect on the specific heat of the salt at low temperatures. Ashmead (1939) found a specific heat anomaly in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ with a maximum at about 1°K , and Duyckaerts (1942) has measured the specific heat at temperatures just above this. Duyckaert's values lead to an excess specific heat above the Debye term of about $1.5/T^2 \text{ cal./mol.}$ at temperatures reasonably above the temperature of the specific heat maximum. If this is attributed to exchange effects, an approximate relation between the excess specific heat and the exchange energy is given by $\frac{C_v}{R} = \frac{3}{4}f \left(\frac{J}{kT} \right)^2$, where $2f$ is the number of nearest neighbours and $2J$ the exchange energy. Using $C_v = 1.5/T^2$ as above, the exchange frequency becomes $1.4/\sqrt{f} \text{ cm.}^{-1}$. For any reasonable value of f this is rather larger than the other values given above.

Again, if the static susceptibility of copper sulphate can be expressed as $\chi = \frac{C}{T + \theta}$, the Curie temperature θ should be affected by the exchange interaction. Measurements on powdered $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ by de Haas & Gorter (1930) down to 14°K and by Reekie (1939) down to 1.6°K both give $\theta = 0.7^\circ\text{K}$. Recent measurements by Benzie & Cooke on a single crystal in the temperature range 10° to 1°K show that the crystal is isotropic so far as the value of θ is concerned. They find a value of $0.6 \pm 0.05^\circ\text{K}$. If this is entirely attributed to exchange interaction, the approximate relation $k\theta = fJ$ gives an exchange frequency of $0.9/f\text{ cm.}^{-1}$, again a rather large value.

These estimates of the exchange frequency are not very consistent although they are of the same order of magnitude. The reasons for the differences may lie in one or more of the following considerations:

- (1) The estimates are based on approximate relations.
- (2) The exchange interaction between similar and dissimilar ions may be different. The value of 0.15 cm.^{-1} is for the exchange between dissimilar ions, and if the interaction between similar ions is greater than this, the average value determined from the specific heat and susceptibility measurements would also be greater.
- (3) The width of the narrowest absorption line may be partly due to other causes, such as the spin-lattice relaxation time being short. The estimate of 0.08 cm.^{-1} obtained from this must therefore be regarded as a lower limit.
- (4) The exchange interaction may be a function of the temperature.

Thus there appears to be need for further experimental work by all these methods to clear up the many points still outstanding.

Our thanks are due to Professor Lord Cherwell for extending to us the facilities of the Laboratory, to many of our colleagues for very helpful discussions, in particular Professor M. H. L. Pryce and Dr B. Bleaney and to the Board of Admiralty for apparatus and other assistance in carrying out the work. One of us (D. M. S. B.) was assisted during the course of these experiments by a grant from the D.S.I.R.

REFERENCES

- Arnold, R. D. & Kip, A. F. 1949 *Phys. Rev.* **75**, 1199.
 Ashmead, J. 1939 *Nature*, **143**, 853.
 Bagguley, D. M. S. & Griffiths, J. H. E. 1947 *Nature*, **160**, 532.
 Bagguley, D. M. S. & Griffiths, J. H. E. 1948 *Nature*, **162**, 538.
 Bagguley, D. M. S., Bleaney, B., Griffiths, J. H. E., Penrose, R. P. & Plumpton, B. I. 1948 *Proc. Phys. Soc.* **61**, 542, 551.
 Beevers, C. A. & Lipson, H. 1934 *Proc. Roy. Soc. A*, **146**, 570.
 Benzie, R. & Cooke, A. H. Private communication.
 Bleaney, B. 1949 *Phys. Rev.* **75**, 1962.
 Bleaney, B., Penrose, R. P. & Plumpton, B. I. 1949 *Proc. Roy. Soc. A*, **198**, 406.
 Bloembergen, N., Purcell, E. M. & Pound, R. V. 1948 *Phys. Rev.* **73**, 679.
 de Haas, W. J. & Gorter, C. J. 1930 *Commun. Phys. Lab. Univ. Leiden*, 210d.
 Duyckaerts, G. 1942 Thesis, Liège.
 Gorter, C. J. 1947 *Paramagnetic relaxation*, p. 67.
 Gorter, C. J. & Van Vleck, J. H. 1947 *Phys. Rev.* **72**, 1128.
 Halliday, D. & Wheatley, J. 1948 *Phys. Rev.* **74**, 1712.

- Jordahl, O. M. 1934 *Phys. Rev.* **45**, 87.
 Krishnan, K. S. & Mookherji, A. 1936 *Phys. Rev.* **50**, 860.
 Krishnan, K. S. & Mookherji, A. 1938 *Phys. Rev.* **54**, 533 and 841.
 Polder, D. 1942 *Physica*, **9**, 709.
 Pryce, M. H. L. 1948 *Nature*, **162**, 538.
 Pryce, M. H. L. & Stevens, K. W. H. 1949 *Proc. Phys. Soc.* **63**, 36.
 Reekie, J. 1939 *Proc. Roy. Soc. A*, **173**, 367.
 Van Vleck, J. H. 1948 *Phys. Rev.* **74**, 1168.
 Wheatley, J. & Halliday, O. 1949 *Phys. Rev.* **75**, 1412.
 Whitmer, C. A., Weidner, R. T., Hsiang, J. S. & Weiss, P. R. 1948 *Phys. Rev.* **74**, 1478.
 Zavoisky, E. 1945 *J. Phys. U.S.S.R.* **9**, 211.
 Zavoisky, E. 1946 *J. Phys. U.S.S.R.* **10**, 197.

The adsorption of vapours on mercury

IV. Surface potentials and chemisorption

BY C. KEMBALL, *Department of Colloid Science, University of Cambridge*

(Communicated by E. K. Rideal, F.R.S.—Received 9 November 1949—
 Read 2 March 1950)

An apparatus has been constructed to measure the surface potential of adsorbed vapours on mercury. There was shown to be a change in potential of 0.055 V at the phase change associated with the alteration in the orientation of toluene molecules adsorbed on the surface of the mercury.

Carbon tetrachloride, hexachlorethane and chloroform vapours reacted with mercury, and the rate of the reaction was determined by the accompanying change in surface potential which was over 1 V in the case of carbon tetrachloride. The kinetics indicated that the substances were dimerizing under the action of the mercury surface with negligible activation energy. The rate of reaction with carbon tetrachloride was in agreement with the calculated rate, assuming that the activated complex consisted of two physically adsorbed molecules loosely held together and able to move over the surface.

Methyl iodide vapour was found to react with mercury only when illuminated with light from a mercury lamp. There was little change of surface potential associated with this reaction, but the kinetics were determined from the irreversible changes in the surface tension of the mercury. It was found that the rate of reaction depended on the square root of the light intensity, which suggested that the methyl iodide was being dissociated into radicals, which in turn reacted with the mercury.

INTRODUCTION

The measurement of surface potentials has proved a useful method of studying insoluble films at liquid surfaces, and it can be extended to the study of the films formed by vapours on metal surfaces as was done by Frost (1942). Two electrodes are necessary to measure the potential, and the vapour will be adsorbed on both, rendering the interpretation of the results difficult when only physical adsorption occurs. However, if a phase change occurs in the adsorbed film on one of the electrodes it may be accompanied by a change in surface potential, the measurement of which will lead to a greater understanding of the nature of the phase change. Likewise, if a chemical reaction occurs slowly at one of the surfaces there may be a drift in the value of the potential which can be correlated with the reaction, and such a

change in potential will remain even after the removal of the vapour. This technique has been used for studying the chemisorption of gases on metals, particularly tungsten, by Bosworth & Rideal (1937) and others. The vapours of carbon tetrachloride and other halogen compounds are known to react with mercury, and the present paper describes an attempt to follow the kinetics of these reactions by the study of surface potentials.

EXPERIMENTAL

The apparatus, which was based on the type used by Kemball (1946), is shown in figure 1. Clean mercury from a still entered at *A*, and after spilling over the edge of the cup *C* the mercury returned to the still through the tube *B*. *D* was used both for evacuation of the apparatus and for the admission of vapours. The mercury formed one electrode, and the nickel plate *E* which was attached by a tungsten wire to a block of iron suspended by a fine tungsten spring formed the second electrode. The

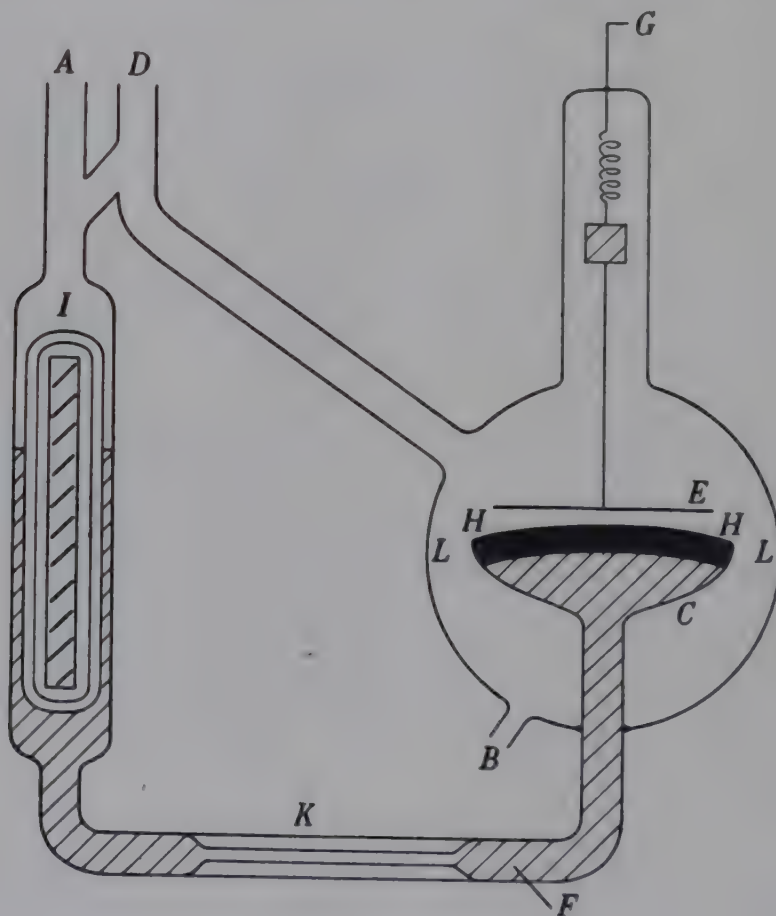


FIGURE 1. Apparatus

position of the nickel plate could be adjusted by means of a permanent magnet acting on the iron block. The two electrical connexions were situated at *F*, which consisted of a tungsten wire sealed through the glass and connected to the potentiometer of the usual type of circuit for the measurement of contact potentials, and at *G* which was connected by a short wire to a Lindemann electrometer. A small projection on the underside of the nickel plate replaced the tungsten pointer on the

earlier apparatus and enabled detection of the top of mercury meniscus *H*. The level of the meniscus could be lowered by lifting the float *I* (iron sealed in glass) by an electromagnet. Oscillations of the mercury were damped by the capillary constriction *K*.

The method employed by Kelvin (1898) was used to measure the contact potential between *H* and *E*. This method depended on the fact that the potential of one plate of a condenser, i.e. *E*, is independent of the capacity of the condenser and hence the position of the meniscus *H*, only when the condenser was uncharged. This condition was obtained by altering the potentiometer so that no movement of the unearthed Lindemann needle occurred on lifting the float *I*. The apparatus was screened, and in order to have the greatest sensitivity, which depended on keeping the capacity of the electrometer and connexions to the minimum, a wide cylinder of gauze was used to screen the short connexion from *G* to the electrometer. It was possible to measure the contact potential to an accuracy of 10 mV. The lip of the cup *C* had to be made conducting by evaporating tungsten on to it, because before this had been done the movement of the mercury away from the glass left the cup charged to a high potential which was estimated to be as much as 200 V and caused a great deflexion of the electrometer needle as the edge of the cup was bared. The process of distillation which was used to clean the mercury resulted in the presence of large charges on the walls of the apparatus, and the act of lowering the mercury meniscus exposed the nickel plate to the influence of these charges, particularly those at *L*. These resulted in a double kick of the electrometer during the lowering of the meniscus which moved in all about 3 mm. The first part of this kick was due to the change in the capacity of the condenser between the plate and the mercury, and the direction of this kick was dependent on the potential supplied by the potentiometer. The second part of the kick was due to the change in the capacity of the condenser formed by the plate and the walls at *L*, and it was not influenced by the setting of the potentiometer. These surface charges rapidly disappeared when water vapour was admitted to the apparatus, because it increased the surface conductivity of the glass. It was necessary to discharge the apparatus in this way each day, and the water vapour was then removed by evacuation for half an hour before conducting an experiment.

Initially an attempt had been made to measure the contact potential by ionizing the vapour between a smaller nickel plate and the mercury by means of mesothorium suspended in a glass capsule attached to the tungsten wire. This method was only successful with upwards of 2 cm. of gas or vapour present. At lower pressures the conductivity was insufficient to conduct away the positive charge resulting from the emission of the β -particles, and the Lindemann needle rapidly assumed a large positive potential.

The methods of purification and the data for the determination of the vapour pressure from the temperature of the reservoir supplying the substances to be adsorbed were as follows:

Toluene. The sample was the same as in the previous work (Kemball & Rideal 1946).

Carbon tetrachloride. The carbon tetrachloride was freed from carbon disulphide by shaking with potassium hydroxide, and after washing with water it was dried

over potassium hydroxide and distilled over mercury. Vapour pressures were determined from the work of Mundel (1913) and Young (1891).

Hexachlorethane. A commercial sample was recrystallized from alcohol and then sublimed *in vacuo*. The results of Ivin & Dainton (1947) were used to obtain vapour pressures.

Chloroform. The sample had been washed several times with water to remove alcohol and dried over calcium chloride before distillation. The refractive index at 29°C was 1.4404, comparing with the expected value of 1.4406. Pressure data were obtained from the work of Drucker & Kangro (1915).

Methyl iodide. This substance was purified by shaking with mercury and then by distillation. The results of Rex (1906) were used to calculate pressures.

With toluene, experiments were performed to measure the surface potential associated with the phase change which had been found to occur at surface pressures of 22 dynes/cm. at 50°C and at 51 dynes/cm. at 25°C (Kemball & Rideal 1946). The results are shown in figure 2, where the contact potential between the nickel and the mercury is plotted against the surface pressure of the adsorbed toluene at the two temperatures. In all the experiments to be described the mercury had been freed from dirt by distillation and evacuation. At both temperatures a change in potential of about -0.055 V was observed at surface pressures corresponding to the phase change. The sign of the potential represents that applied to the mercury, and a negative sign means that there was a decrease in the work function of the mercury.

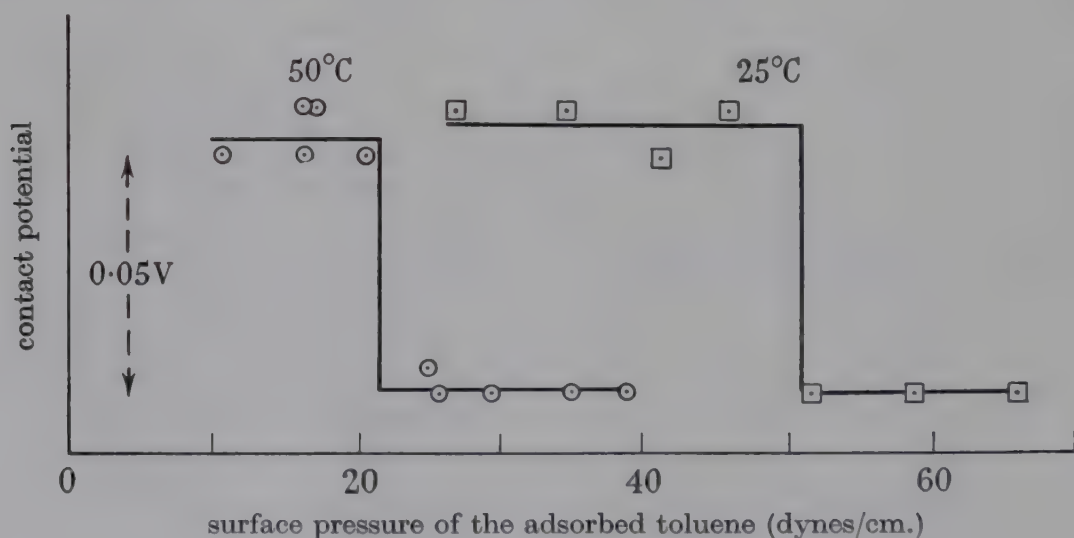


FIGURE 2. Surface potential of toluene films.

The three substances carbon tetrachloride, hexachlorethane and chloroform all gave experimental results of the same type. When the vapour was admitted to clean mercury the surface potential altered at a rate which varied with the pressure of the gas and which gradually diminished after a few minutes. The results of several experiments using different pressures of these three substances at two temperatures are shown in figures 3 to 8. In order to ensure that the required pressure of vapour was attained throughout the apparatus as quickly as possible the reservoir supplying the vapour was allowed to fill a large bulb, before the cut-off to the mercury was

lowered. Removal of the carbon tetrachloride vapour at any stage of an experiment brought the change of surface potential to a standstill, and the potential would revert to the original value after the mercury drop had been spilt several times by the distillation of more mercury, showing that the drift of surface potential was clearly to be associated with chemisorption of the vapour on the mercury surface. When the vapour of hexachlorethane was removed during an experiment the alteration of the surface potential gradually ceased. This was owing to the time taken to evacuate to a sufficiently small pressure to prevent further reaction. Removal of chloroform vapour led to some reversal of the change in the surface potential. This was attributed to the lower stability of the film formed by this substance, and it was thought that the movement of the mercury which was necessary in the measurement of the surface potential was sufficient to 'break' the chemisorbed film. Prolonged agitation of the surface also led to some reduction in the potential of the chemisorbed film from carbon tetrachloride. In all experiments a greater positive potential had to be applied to the mercury as the reaction proceeded, showing that there was a gradual increase of the work function of mercury.

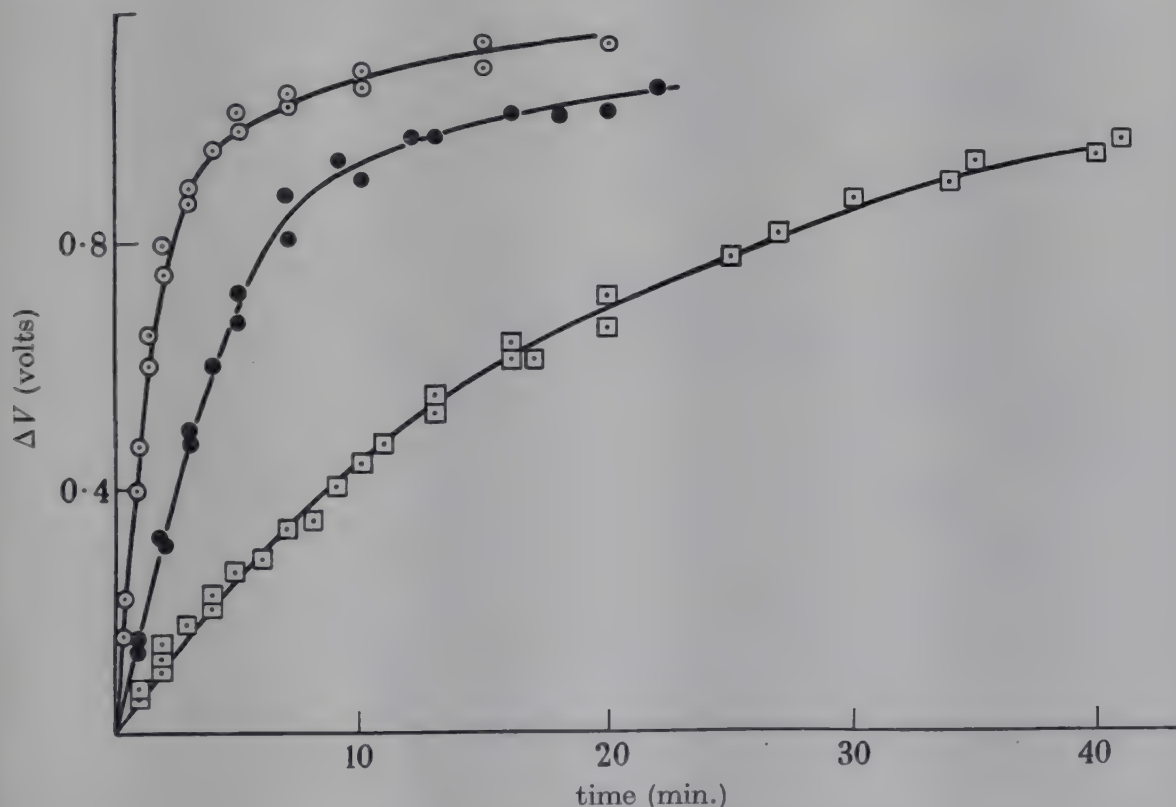


FIGURE 3. Reaction of carbon tetrachloride with mercury at 25° C.
Pressures: $\odot$, 0.21 mm.; $\bullet$, 0.123 mm.; $\square$, 0.074 mm.

In the case of carbon tetrachloride an attempt was made to determine the relationship between the change in the surface potential and the irreversible surface pressure by removal of the vapour after different periods of contact and measuring the surface tension of the mercury with the chemisorbed film. The results are shown in figure 9a.

No reaction occurred when methyl iodide vapour was admitted to mercury in the absence of light of wave-length less than about 3400 Å. The admission of 25 mm. of vapour caused a change in the surface potential of -0.52 V (in the opposite sense

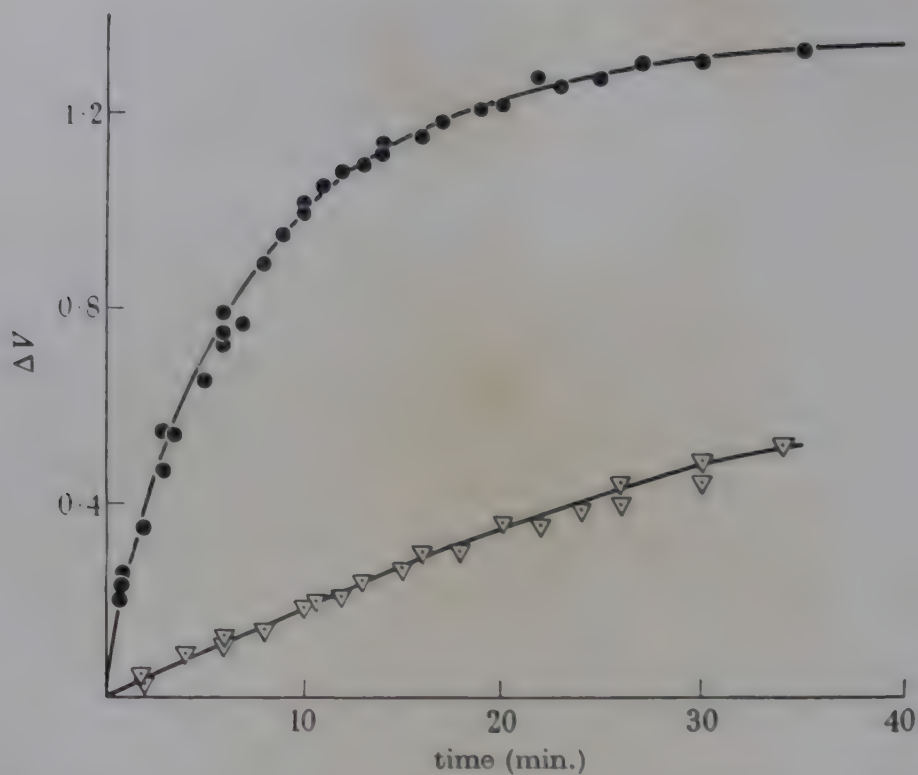


FIGURE 4. Reaction of carbon tetrachloride with mercury at 50° C.
 ●, 0.123 mm.; ▽, 0.043 mm.

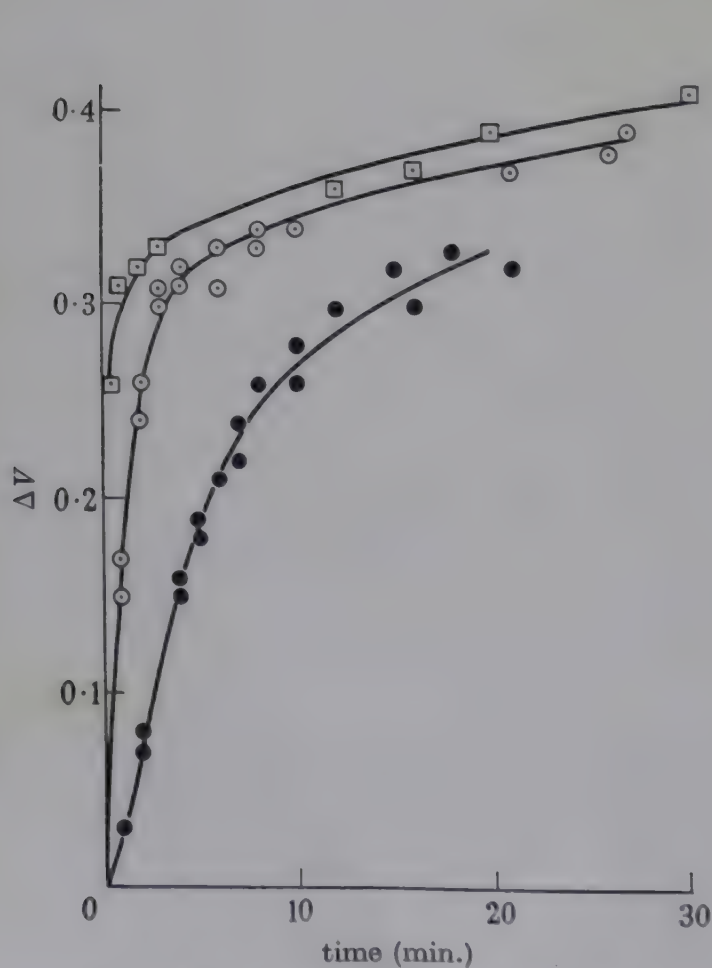


FIGURE 5. Reaction of hexachlorethane with mercury at 25° C. □, 0.077 mm.;
 ○, 0.042 mm.; ●, 0.022 mm.

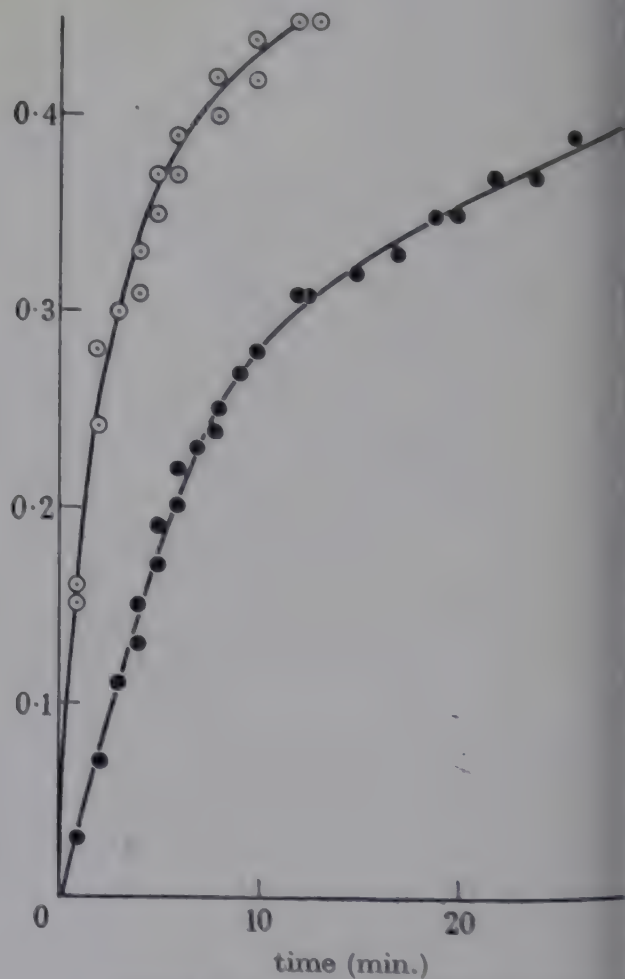


FIGURE 6. Reaction of hexachlorethane with mercury at 50° C. ○, 0.042 mm.;
 ●, 0.022 mm.

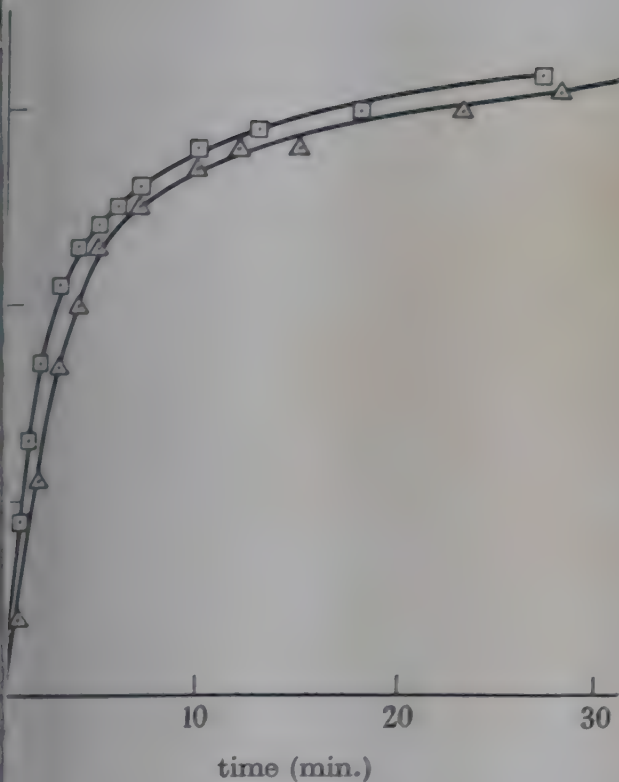


FIGURE 7. Reaction of chloroform with mercury at 25° C. □, 2.88 mm.; Δ, 2.08 mm.

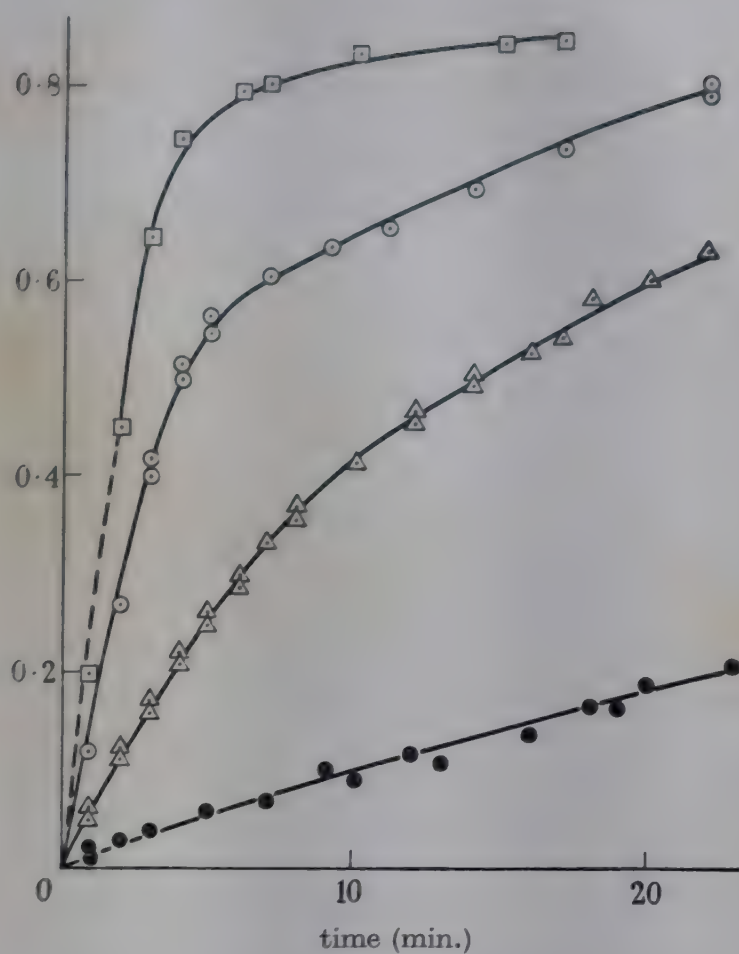


FIGURE 8. Reaction of chloroform with mercury at 50° C. □, 17.9 mm.; ○, 4.32 mm.; Δ, 2.08 mm.; ●, 0.91 mm.

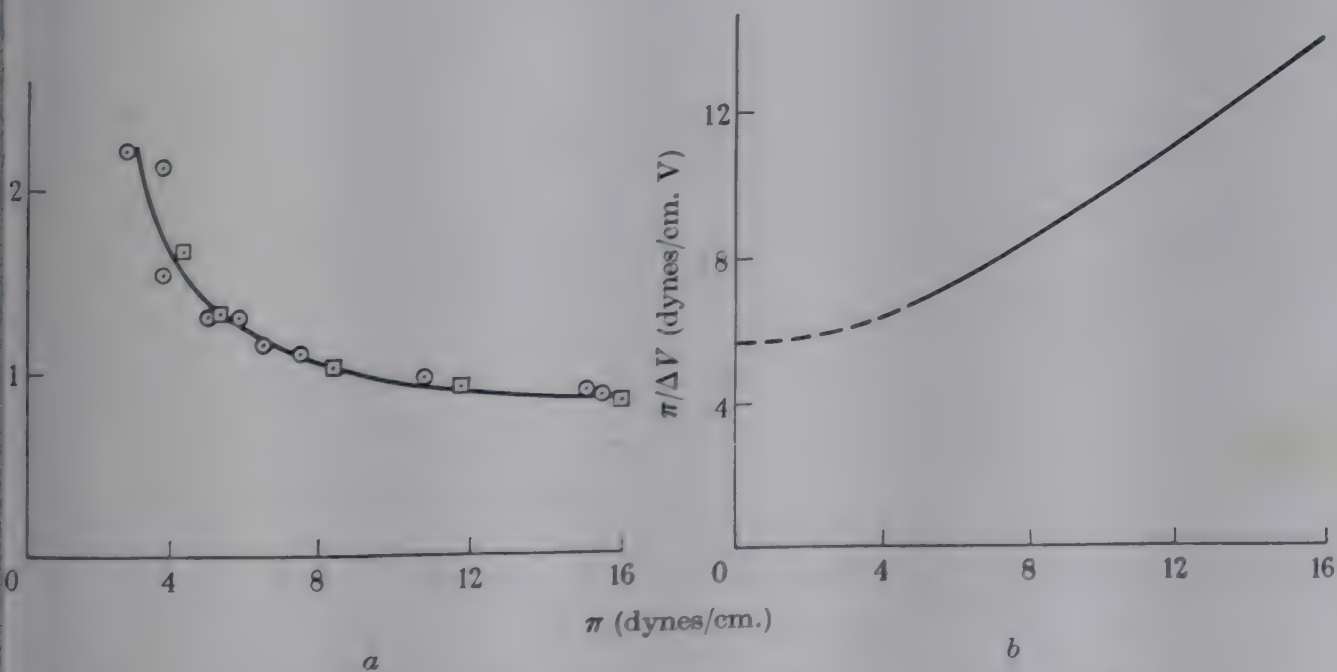


FIGURE 9. Chemisorbed carbon tetrachloride on mercury. ○, 25° C.; □, 50° C.

to the changes with the chlorine compounds) and a surface pressure of about 60 dynes/cm., but on removal of the vapour the surface potential disappeared and the surface tension reverted to within a dyne/cm. of its original value. However, when the vessel was illuminated with light from a mercury lamp for 10 min. the contact potential altered from -0.52 to -0.48 , i.e. a change of $+0.04$ V, and on removal of the vapour there was found to be an irreversible surface pressure of some 38 dynes/cm., and the surface potential had altered by $+0.05$ V from the original value before the vapour had been admitted. Unfortunately, the surface potential associated with this photochemical reaction was not sufficiently great to be used to follow the kinetics of the reaction, and it was necessary to measure the irreversible surface pressures instead.

It was found that ordinary electric light and the lines 3650 and 3663 (obtained from the mercury lamp by using a filter of 3 mm. of Chance's black glass and 2 cm. of a solution of 62.5 g./l. of hydrated copper sulphate) produced no reaction. The Pyrex glass of the apparatus would have stopped all light of wave-length less than 3000 Å, and therefore the reaction must have been due to the 3341 or 3125 lines of the mercury lamp. A blank test showed that illumination caused no change in the surface tension of mercury in the absence of vapour.

A series of experiments were carried out using 17.4 mm. of vapour and measuring the irreversible surface pressure after the vapour had been illuminated for different periods. The results are given in table 1. Another series of experiments using different pressures and different intensities of light (obtained by means of calibrated gauzes) but illumination for a fixed period gave the results shown in table 2.

TABLE 1. PHOTOCHEMICAL REACTION OF METHYL IODIDE WITH MERCURY

		pressure 17.4 mm.							
time of illumination (min.)	0	10	21	30	42	50	60	70	80
irreversible surface pressure (dynes/cm.)	0	23.9	39.9	50.2	58.4	63.5	67.2	70.0	70.7

TABLE 2. PHOTOCHEMICAL REACTION OF METHYL IODIDE WITH MERCURY

		illumination for 10 min.		
		irreversible surface pressure (dynes/cm.)		
vapour pressure	...	6.0 mm.	17.4 mm.	25.1 mm.
light intensity				
11		—	4.4	15.4
23		—	8.0	22.3
52		3.7	13.7	31.1
100		6.7	18.7	38.2

INTERPRETATION AND DISCUSSION

The results of the experiments with toluene confirmed the existence of the phase change in the adsorbed film found in the earlier work (Kemball & Rideal 1946). It was suggested that this represented the transition from a gaseous film with the ring of the toluene parallel to the surface having a co-area of 37 Å² to a gaseous film with

the ring perpendicular to the surface and having a co-area of $23 \text{ \AA}^2$. The change in the potential of -0.055 V indicated that a slight reduction in the work function of the mercury was associated with alteration in the orientation of the toluene molecules. A reduction in work functions would be caused by dipoles with the negative end orientated towards the mercury and toluene has a small dipole moment of 0.43 debye units directed from the methyl group towards the ring, indicating that in the second phase the molecules are orientated with the ring towards the mercury.

The results with carbon tetrachloride, hexachlorethane and chloroform may be treated as a group. All gave changing potential with time because of the reaction of the vapour with the mercury surface. Reaction of the vapour with the nickel will not explain the results, because the surface of the nickel remained the same throughout all the experiments, whereas a freshly prepared surface of mercury was used in each test. The nickel surface may have undergone reaction with the vapours in initial trial experiments, but thereafter would have remained unchanged. In all cases the reaction caused a change of potential with time which was initially linear, and gradually decreased, presumably due to saturation of the surface. The sign of the change indicated that the work function of the mercury was increasing, indicating the development of dipole moments directed away from the mercury. The dipole moments which affected the work function in these cases of chemisorption must have been located in the mercury-adsorbate bond (not above the mercury as in the case of the physically adsorbed toluene) and were presumably associated with Hg-Cl bonds.

A plot of the logarithm of the initial rates of change of surface potential against the logarithm of the pressures for each substance is shown in figure 10. The gradients of the lines obtained are approximately 2 at low pressures, indicating that the initial reactions were second order with respect to the pressure of the vapours and therefore involved two molecules of each substance. The departure from second-order behaviour which occurred with chloroform at higher pressures would be expected if the reactants were physically adsorbed molecules, the number of these becoming less dependent on the pressure as the surface was filled up. It was not possible to obtain accurate results for the other two substances at higher pressures because the reactions were too rapid to be measured by this technique.

If the change in surface potential corresponded to the amount of chemisorbed material, the graphs of the measurements with carbon tetrachloride of surface pressure against the reciprocal of the surface potential, i.e. π against $1/\Delta V$ and of $\pi/\Delta V$ against π (figure 9*a* and *b*) correspond to the plots of π against A and πA against π , which are often used to describe adsorbed films, A being the area per molecule. The slope of the higher portion of the second graph indicates that as the value of the surface pressure increased, the value of ΔV tended to 1.45 V , which would therefore be the surface potential of a complete film of chemisorbed carbon tetrachloride. The fact that the initial portion of the second graph is approximately horizontal may indicate that at low-surface pressures the film obeyed the relationship $\pi A = kT$ which is often found to be true for gaseous films at low pressures. The fact that the film in this case consisted of chemisorbed units does not mean that it could not be gaseous because it would be possible for each chemisorbed unit with its

underlying mercury atom or atoms to move freely about on the surface. If the value of $\pi/\Delta V$ at zero pressure is assumed to correspond to the value of kT at 310°C which may be expressed as $412 \text{ \AA}^2 \text{ dynes/cm.}$, the gradient of the portion of the curve at higher pressures corresponds to a value of $50 \text{ \AA}^2$, which is a reasonable value for the size of the chemisorbed unit. The extrapolation to zero surface pressure cannot be performed accurately, and therefore the value of $50 \text{ \AA}^2$ is only approximate.

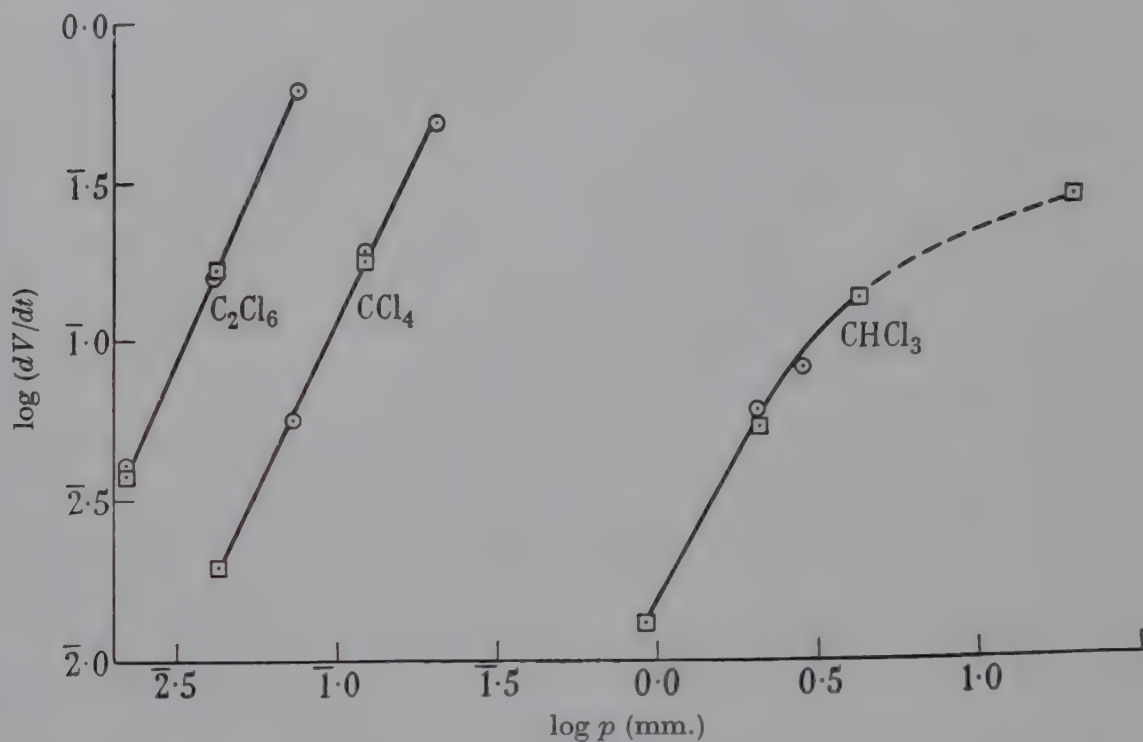


FIGURE 10. The effect of pressure on the initial rates of reaction. ○, 25°C .; □, 50°C .

Comparison of the rates of reaction at 25 and at 50°C for the three substances (figure 10) indicates that the apparent energy of activation of the initial reaction is zero with respect to the energy of a gas molecule in each case. The experiments with 0.123 mm. of carbon tetrachloride at the two temperatures shown in figures 3 and 4 indicate that the rates of reaction were the same until the surface potential reached 0.7 V , and at higher surface potential the higher temperature gave a faster rate. The apparent energies of activation throughout the course of the reaction were calculated by comparing the slopes of the curves at corresponding values of surface pressure and they are given in table 3. Since with a completely covered surface the surface potential would have been 1.45 V , the calculations show that the energy of activation started to increase as soon as the surface was half covered. This type of behaviour is similar to the variation of the heat of adsorption of a mobile film when there is interaction between the adsorbed molecules as calculated by Wang (1937). With less than half the surface covered there is little decrease in the heat of adsorption, as no molecule need approach near its neighbours, but once that stage is passed the interaction may be considerable. Similar considerations may have applied to the adsorption of the carbon tetrachloride molecules because of the strong dipole repulsion that would exist between the chemisorbed units and the activated complex.

Inspection of figures 5 to 8 shows that there was an energy of activation for the reaction of hexachlorethane after the surface potential reached 0.26 V, and in the case of the chloroform reaction after 0.23 V.

TABLE 3. ENERGY OF ACTIVATION FOR THE REACTION
OF CARBON TETRACHLORIDE WITH MERCURY

surface potential (V)	0.0	0.2	0.4	0.6	0.8	0.9	1.0	1.1
activation energy (kcal./g. mole)	-0.1	-0.3	-0.5	-0.4	1.7	4.1	9.6	19.2

It is believed that these three substances were each undergoing dimerization on the surface of the mercury. It is known that hexachlorethane can be formed from carbon tetrachloride by the action of a number of metals, e.g. Cu, Ag, Al and Hg, which are converted to their chlorides. The hexachlorethane formed by the reaction of carbon tetrachloride will in turn react to form higher chlorinated hydrocarbons, and the reaction once started might have proceeded indefinitely, but for the fact that the free surface was rapidly becoming blocked with reaction products. The fact that hexachlorethane did not give such large surface potentials may have been due to the blocking of the surface by the large molecules of chlorinated butane formed which would be very strongly adsorbed. The smaller potentials with chloroform may have been due to the product of the initial dimerization, i.e. the tetrachlorethane being less reactive.

It is presumed that after reaction the surface was covered with chemisorbed chlorine and such physically adsorbed chlorinated hydrocarbons as would undergo no further reaction. The stability of the film may have depended on the presence of these physically adsorbed components preventing the aggregation of the mercury chloride. The reversal of the potential change observed on evacuation of chloroform vapour may have been due to the removal of the tetrachlorethane (assuming this to be sufficiently unreactive to undergo further dimerization), followed by aggregation of the chlorinated mercury atoms. It must be remembered that the measurement of the potential involved the movement of the surface, which would be a further factor in destroying the nature of the film.

Since the apparent energy of activation of these reactions was zero, it is perhaps surprising that the reactions were slow enough to be measurable. Some confirmation of interpretation of the experiments may be obtained by estimating the theoretical rate of reaction according to the method of the transition state theory as outlined by Glasstone, Laidler & Eyring (1941). From figure 10 the initial rate of reaction of carbon tetrachloride at a pressure of 0.1 mm. corresponded to a change of surface potential of 0.11 V/min., and assuming that full coverage of the surface would have given a potential of 1.45 V and that each chemisorbed unit occupied $50 \text{ \AA}^2$ the rate of reaction was 2.5×10^{11} chemisorbed units/sec.cm.². It is postulated that the activated complex consisted of two physically adsorbed molecules loosely held together so that each retained its normal rotational and vibrational degrees of freedom. The complex would have only two degrees of translational freedom, both in the plane of the mercury surface. The four lost translational degrees of freedom would be replaced by a rotation of the complex as a whole in the plane of the surface,

and three vibrations; one between each molecule and the mercury, and one between the two molecules. One of these three vibrational modes would, however, be adsorbed in the so-called reaction co-ordinate of the activated complex. The rate of reaction would be given by

$$v = c_g^2 \frac{kT}{h} \frac{F_*}{F_g^2} e^{-\epsilon/kT},$$

where c_g represents the number of molecules per ml. of gas at 0.1 mm. pressure, ϵ the energy of activation with respect to the energy of a molecule in the gas phase (known experimentally to be zero), F_g is the partition function for unit volume of vapour and F_* is the partition function for unit area of activated complex (cf. Glasstone, Laidler & Eyring 1941, chap. VII, equation 15). The partition function for the gas is given by

$$F_g = \frac{(2\pi mkT)^{\frac{3}{2}}}{h^3} \left[\frac{1}{\pi\sigma} \left(\frac{8\pi^3 I_A kT}{h^2} \right)^{\frac{1}{2}} \right] f_v^9,$$

where I_A is the moment of inertia of the carbon tetrachloride, σ is the symmetry number ($=12$) and f_v represents a vibrational partition function. The function for the activated complex may be expressed by

$$F_* = \frac{2\pi 2mkT}{h^2} \left[\frac{1}{\pi\sigma'} \left(\frac{8\pi^3 (I_A^6 I_B)^{\frac{1}{2}} kT}{h^2} \right)^{\frac{1}{2}} \right] f_v^{18} f_w^2,$$

where I_B is the moment of inertia for the rotation of the complex as a whole, $2m$ represents the mass, σ' is the new symmetry number ($=12^2 \times 2 = 288$) and f_w represents one of the new vibrational modes. It will be observed that the complex has 29 degrees of freedom, whereas each of the molecules that entered the complex had 15, the remaining degree having been absorbed in the reaction co-ordinate. It is possible to reduce the ratio of the partition functions to

$$\frac{F_*}{F_g^2} = \frac{h^3}{m^2} \left(\frac{I_B \pi}{2(kT)^3} \right)^{\frac{1}{2}} f_w^2,$$

and using the value of 4.9×10^{-37} g.cm.² for I_B the reaction rate is given by

$$v = 2.1 \times 10^{10} f_w^2 \text{ chemisorbed products per sec.cm.}^2.$$

It is not possible to estimate the value of f_w with any accuracy, but if the vibrations, which will be relatively slow as they are associated with physical interaction and not chemical bonds, are assumed to have a frequency of 3×10^{12} sec.⁻¹, the value of f_w will be approximately 2. The calculated rate is thus 8×10^{10} chemisorbed units per sec.cm., and it compares remarkably well with the observed value of 2.5×10^{11} and confirms the supposition that the reaction takes place through the close approach of two physically adsorbed molecules of carbon tetrachloride, the pair being in contact with the mercury but able to move about the surface.

The experiments with methyl iodide showed that reaction of the vapour with the mercury only occurred when light of wave-length less than 3600 Å was used. Iredale & Mills (1931) showed that methyl iodide vapour has an absorption band terminating about 3200 Å. They found that the wave-length of the band-edge varied with the pressure ranging from 3010 Å at 7.5 mm. to 3360 Å at 264 mm. when using a column

of 100 cm. of vapour. This indicates that the 3125 line of the mercury lamp would have caused dissociation of the iodide and that the 3341 line may have had some effect at the higher pressures.

The change in surface potential may have been small because there was simultaneous reaction on the nickel as well as on the mercury surface, but it is more likely that the reaction was not greatly affecting the work function of the mercury. Because of the smallness of the change in surface potential it was necessary to rely on measurement of the surface pressure of the film to determine the kinetics of the photochemical reaction. The data given in table 1 may be interpreted on the assumptions that complete coverage of the surface corresponded to a surface pressure π_∞ , and that at any stage the rate of reaction as measured by $d\pi/dt$ was proportional to $(\pi_\infty - \pi)$. This gives the normal first-order kinetic expression

$$\frac{d\pi}{dt} = k(\pi_\infty - \pi),$$

which may be integrated to give

$$-\ln(\pi_\infty - \pi) = kt + c,$$

and a plot of this type is shown in figure 11 using 76 dynes/cm. as the value of π_∞ . Despite the doubtful character of the assumptions it appears that the relationship fits the facts. From table 2 it is clear that the rate of the reaction depended slightly on the intensity of the light and markedly on the pressure of the vapour. If the

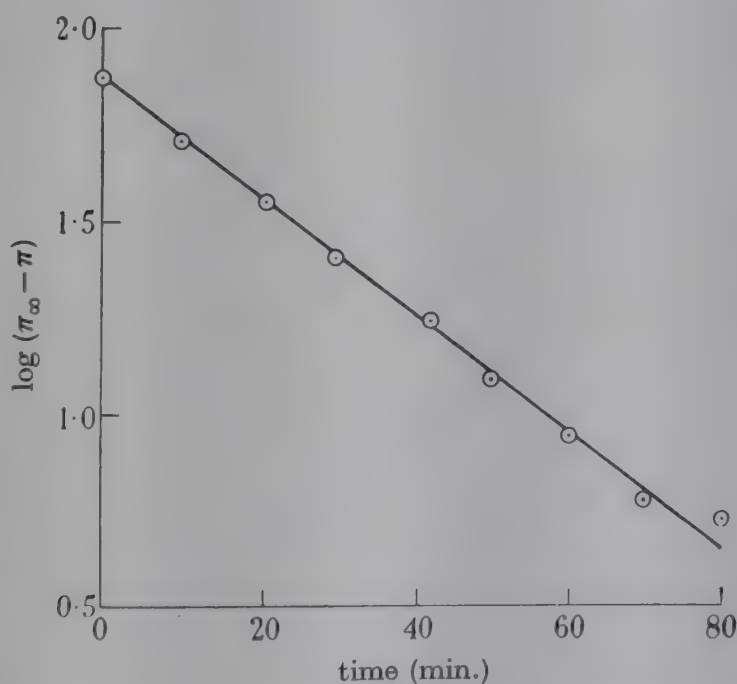


FIGURE 11. Photochemical reaction of methyl iodide with mercury at 25°C.
 $p = 17.4$ mm.; $\pi_\infty = 76$ dynes/cm.

action of the light set up a stationary concentration of radicals in the gas phase which were undergoing re-combination mainly by three-body collision in the gas phase and, if it is assumed that any radicals reaching the surface underwent chemisorption, the experimental data may be explained. The equilibrium concentration

of radicals (which would be initially methyl radicals and iodine atoms, although others may have been formed as well by radical-molecule collisions) may be expressed as

$$k_1 I = k_2 [R]^2 p,$$

where I is the amount of light absorbed, $[R]$ is the concentration of radicals and p is the pressure. I will depend on the incident light intensity and also on some function of the pressure of the vapour, i.e.

$$I = I_0 f(p).$$

The rate of reaction will depend on the concentration of radicals and will be given by

$$\frac{d\pi}{dt} = k_3 [R] (\pi_\infty - \pi),$$

and hence

$$\frac{d\pi}{dt} = k' \left(\frac{I_0 f(p)}{p} \right)^{\frac{1}{2}} (\pi_\infty - \pi),$$

or in the integrated form by

$$-\ln (\pi_\infty - \pi) = k' \left(\frac{I_0 f(p)}{p} \right)^{\frac{1}{2}} t + c'.$$

The relationship between π and t in this equation has already been tested by the plot shown in figure 11 of results obtained at constant values of I_0 and p . The results in table 2 may be used to check the relationship between π and I_0 by plotting $\log (\pi_\infty - \pi)$ against $I_0^{\frac{1}{2}}$, because the value of t was constant in each of these experiments and the straight lines at each pressure are shown in figure 12. The relative slopes of the three lines shows the variation of $(f(p)/p)^{\frac{1}{2}}$ with p . The figures in arbitrary units are 36, 113 and 307 for pressures of 6.0, 17.4 and 25.1 mm., showing that $f(p)$ is

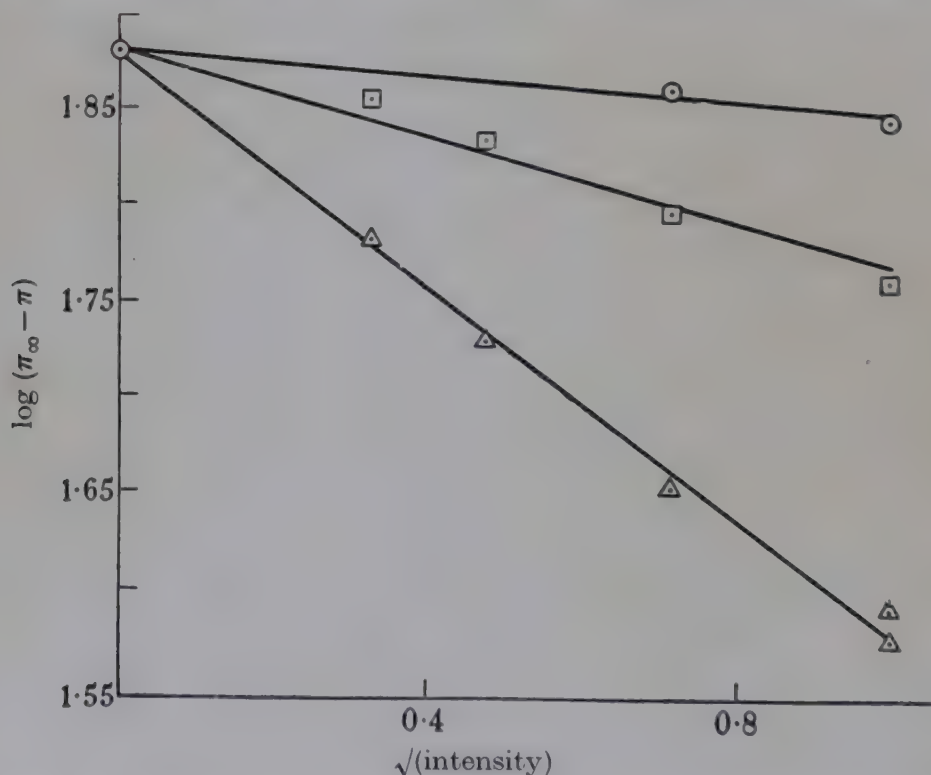


FIGURE 12. Photochemical reaction of methyl iodide with mercury at 25° C
Time of reaction, 10 min.; ○, 6.0 mm.; □, 17.4 mm.; △, 25.1 mm.

strongly dependent on p , as would be expected from the work of Iredale & Mills where the band-edge was found to move to longer wave-lengths with increase of pressure.

This dependence of rate of reaction on the square root of incident light intensity and the fact that light of longer wave-length produced no effect, confirm the suggestion that the reaction was due to dissociation of the iodide followed by chemisorption of the radicals as they reached the surface, the majority of the radicals, however, re-combining by three-body collisions in the gas phase. It is not possible to say whether only the iodine atoms reacted with the mercury or whether the methyl and other radicals formed by radical-molecule interactions were also capable of combination with the mercury.

REFERENCES

- Bosworth, R. C. L. & Rideal, E. K. 1937 *Proc. Camb. Phil. Soc.* **33**, 394.
Drucker, C. & Kangro, W. 1915 *Z. phys. Chem.* **90**, 518.
Frost, A. A. 1942 *Trans. Electrochem. Soc.* **82**, 259.
Glasstone, S., Laidler, K. J. & Eyring, H. 1941 *The theory of rate processes*.
Princeton University Press.
Iredale, T. & Mills A. G. 1931 *Proc. Roy. Soc. A*, **133**, 430.
Ivin, K. J. & Dainton, F. S. 1947 *Trans. Faraday Soc.* **43**, 32.
Kemball, C. 1946 *Trans. Faraday Soc.* **42**, 526.
Kemball, C. & Rideal, E. K. 1946 *Proc. Roy. Soc. A*, **187**, 53.
Kelvin, Lord 1898 *Phil. Mag.* **46**, 91.
Mundel, C. F. 1913 *Z. phys. Chem.* **85**, 454.
Rex, A. 1906 *Z. phys. Chem.* **55**, 358.
Wang, J. S. 1937 *Proc. Roy. Soc. A*, **161**, 127.
Young, S. 1891 *J. Chem. Soc.* **59**, 911.

The permeability of porous materials

By E. C. CHILDS, SC.D. AND N. COLLIS-GEORGE, PH.D., *School of Agriculture,
University of Cambridge*

(Communicated by Sir Geoffrey Taylor, F.R.S.—Received 16 November 1949—
Read 2 March 1950)

The permeability of a porous material to water is a function of the geometry of the boundary between the solid component and the pore space. Expressions of the Kozeny type purporting to represent this function are based upon the particle size or specific surface of the solids, and whilst, for engineering practice, they have given satisfaction for saturated sands, they may fail badly in other cases. By developing a Kozeny type of expression for the particular structure of a bundle of capillary tubes of assorted radii, we demonstrate the cause of the failure.

Such failure may be avoided by relating permeability to pore-size distribution, which is the factor of prime concern and which may be measured directly by even simpler means than are used to determine particle-size distribution. The pore-size distribution is arrived at by an interpretation of the moisture characteristic of the material, i.e. of the curve of moisture content plotted against pressure deficiency. A simple statistical theory, based upon the calculation of the probability of occurrence of sequences of pairs of pores of all the possible sizes, and of the contribution to the permeability made by each such pair, leads to an expression of the permeability as the sum of a series of terms. By stopping the summation at a selected upper limit of pore size one may calculate the permeability at any chosen moisture content and plot it as a function of that content. An example is presented, using a coarse graded sand specified by its moisture characteristic.

To check these calculations, experimental determinations of the permeabilities of unsaturated materials are presented, using two different grades of sand and a sample of slate dust, the results being compared with computed values. The agreement seems good, and is certainly better than that provided by the Kozeny formula as developed, with difficulty, for the purpose.

The limitations and possible improvements of our concept are very briefly discussed, and finally it is shown how a combined use of the moisture characteristic and the permeability (which is itself derivable from the moisture characteristic) leads to an expression for the coefficient of diffusion of water in the material as a function of moisture content. From this it should be possible, in principle, to calculate in suitable cases the course of water movement down a gradient of moisture content. Such a calculation awaits a satisfactory solution of the problem of non-linear diffusion.

The transmission of fluids by porous bodies has widespread relevance to engineering, geological and agricultural problems. The laws of flow and solutions of particular problems have received considerable attention. It is customary to treat problems as essays in the solution of Laplace's equation, it being assumed, sometimes too readily, that the material of the body obeys Darcy's law, which may be written

$$v = -K \text{ grad } \phi, \quad (1)$$

where v is the flow velocity commonly expressed as ml./sec./unit area normal to v , ϕ is the hydraulic potential, and the constant K , characteristic of the material, is called the permeability of the material to the fluid. This law seems to be valid for Reynolds numbers of less than unity (Fancher, Lewis & Barnes 1933).

The one physical property of the material which enters into a flow problem is the permeability, and this must be known if a complete solution is to be obtained. In principle a measurement of permeability is a simple matter of measuring the rate of

flow of liquid in a column of the material between planes of measured separation and hydraulic potential. In fact, there is considerable divergence of opinion as to the significance of such measurements. Some engineers report them without comment as a matter of routine, whilst other observers have dilated upon the often uncontrollable variations of permeability measured in this way during the course of any single experiment, and have provided a fairly considerable literature in attempting to account for these apparent variations and to indicate an interpretable technique. Permeability has been held to decrease with time owing to the percolating liquid (water) releasing dissolved air into the pores; to swelling of colloidal material; to mechanical blocking by movement of the finest particles of non-cemented materials; to the growth of organisms in the pore space; and to the chemical effect of the flowing liquid upon the material, as, for example, by exchange of bases between the colloidal fraction of the material and saline percolating water, resulting in modification of the colloidal properties of that fraction. Permeability has been held to increase with time due to solution in the percolating water of initially entrapped air. Some of these findings are summarized by Christiansen (1944). In our own experience with low hydraulic heads, the pore space is much modified in the neighbourhood of retaining grids which may be used to support sands in glass tubes, giving rise to anomalously high potential gradients over a portion of the column, whilst manometers measuring the pressure components of such potentials have proved insensitive and capricious. It is therefore of interest to inquire whether the permeability may be inferred from a knowledge of the pore-space geometry by which it is clearly uniquely determined, and which is itself not subject to such experimental uncertainties of determination. There are those who take the argument in reverse and regard a measurement of permeability as the simplest method of inferring the specific surface of a porous material, but we shall show that the argument is not sound.

In the past it has been more common to describe a granular material in terms of particle-size distribution (mechanical composition) than to regard it as a porous material with a given pore-size distribution. As a consequence, current formulas relating permeability to constitution do so in terms of particle size. The best known is perhaps that of Kozeny (1927), but somewhat similar expressions have been derived by Fair & Hatch (1933), Terzaghi (1925) and Zunker (1933); all derive in some degree from Slichter (1899). Kozeny's formula may be given in the form

$$K = Cd^2p^3/(1-p)^2, \quad (2)$$

where p is the porosity (ratio of pore volume to total apparent volume), C is an empirical constant, and d is a measure of the particle size, which is clearly ill-defined in a sand with a wide range of particle sizes. Fair & Hatch substitute for d the factor (V/A) , where A is the total surface of a representative sample of particles the sum of whose volumes is V , that is, A/V is the specific surface. Fair & Hatch's form is therefore

$$K = Dp^3(V/A)^2/(1-p)^2. \quad (3)$$

Zunker & Terzaghi present formulas in which the functions of p are somewhat different from (2) and (3).

Objection may be raised against all of these formulas on various grounds. It is customary for advocates of any one to claim agreement with observed values of permeability, and sometimes superior agreement as compared with others. The fact seems to be that it has been impossible to secure a sufficiently wide range of variation of p and of pore-size distribution to test the formulas severely; and such tests as have been reported indicate a wide range of error (Muskat 1937). That the method of determining the specific surface of powders by measurement of permeability has achieved considerable popularity (Carman 1938, 1939; Lea & Nurse 1939; Rigden 1943, 1947) may reflect only the essential similarity of most industrial powders, and the margin of error which is permissible in the industrial problems involved. As to objections on grounds of principle, it seems to be generally admitted that these formulas all fail utterly to describe structured bodies such as, for example, the 'stiff-fissured' clays. The structural fissures contribute negligibly both to porosity p and to specific surface, and yet they dominate the permeability K . Again, neither p nor (V/A) are directed quantities, and consequently formulas such as (2) and (3) cannot indicate anisotropic permeability, which nevertheless seems to be the rule rather than the exception in nature.

The root of the error in Kozeny-type formulas is readily demonstrated by developing such a formula, by the method of Fair & Hatch, for a hypothetical structure, namely, a bundle of capillary tubes. Consider a tube of radius r conducting a liquid at a rate of δv ml./sec. as a consequence of a potential gradient $d\phi/dl$. By Poiseuille's equation we have

$$\delta v = -(\pi r^4/8\eta)(d\phi/dl),$$

where η is the viscosity of the liquid. If there are n of these tubes in an otherwise solid cross-section of unit area, the total flow is

$$v = -(n\pi r^2/8\eta)(2\pi r^2/2\pi r)^2(d\phi/dl), \quad (4)$$

which may be written

$$v = -(p/2\eta)\{(V/A)p/(1-p)\}^2 d\phi/dl.$$

Comparison with equation (1) yields

$$K = (1/2\eta)p^3(V/A)^2/(1-p)^2, \quad (5)$$

which is Fair & Hatch's form of Kozeny's equation; this equation is therefore demonstrably valid for the specified structure.

If now we consider capillaries of varying sizes, providing an overall porosity p , we may take a group of capillaries, each of radius r , associate with it a proportion of the intervening solid to provide the group with porosity p , thus arriving at a total cross-section a_r of conductor permeated with capillaries of radius r . The permeability contribution K_r of this group enters into the total permeability according to the expression

$$K = \Sigma a_r K_r / \Sigma a_r, \quad (6)$$

in which K_r is given by (5) with a value $(V/A)_r$ proper to the group of radius r . Hence we have

$$K = (1/2\eta)\{p^3/(1-p)^2\}\{\Sigma a_r (V/A)_r^2 / \Sigma a_r\}. \quad (7)$$

The final bracket may be written $(\bar{V}/A)^2$ and represents the mean value of $(V/A)_r^2$ of all the groups; it is clearly not identical with the overall value of $(V/A)^2$ for all channels together, since the former is dominated by the larger and the latter by the smaller channels. Equation (7) is certainly not Kozeny's equation. However, if we consider two similar structures, i.e. structures λ and μ , such that for every group of channels of radius r in λ there is a corresponding group of channels of radius mr in μ , each group contributing the same fraction to the total respective cross-section, then it is easy to show that

$$\overline{(V/A)_\lambda^2}/\overline{(V/A)_\mu^2} = (V/A)_\lambda^2/(V/A)_\mu^2 = 1/m^2. \quad (8)$$

Since, by hypothesis, p is the same for both structures, (7) gives

$$K_\lambda/K_\mu = (V/A)_\lambda^2/(V/A)_\mu^2, \quad (9)$$

and it is only in this restricted sense that Kozeny's equation may be said to be valid for bundles of capillary tubes. We may note that it is necessary to this validity that the porosities of the two structures should be identical, since otherwise the structures cannot be similar. It may be objected that what we have proved for bundles of capillaries may have no relevance to such structures as sand beds, and that must be granted. We would only point out that it has been an accepted procedure to derive Kozeny's equation by a bold development of Poiseuille's (Fair & Hatch 1933; Rigden 1947), without demonstrable justification; it seems to us a far more plausible supposition that the *limitations* we have demonstrated for capillary tubes apply equally to sand beds.

We have sought to relate permeability more rationally with pore-size distribution, which, in recent years, has become almost as common a determination as mechanical composition (Donat 1937; Childs 1940, 1942; Feng & Browning 1946). At the same time we have developed a more stringent test of all such relationships by devising a method of measuring the permeability of a sand column over a wide and controlled range of effective porosity and over a variety of types of pore-size distribution. The latter has been accomplished by measuring the permeability over the complete range of moisture content. A pore which is filled with air is not effective in conducting water; the effective porosity is limited to that which is water-filled. Also, since larger pores lose their water before smaller ones, reduction of moisture content results also in a steady change of pore-size distribution, but this distribution is not, of course, variable at will. In this paper, therefore, we present a comparison of the computed variations of permeability of a number of different packed sands and dusts with moisture content with the variation observed in experiments.

The pore space of a material such as sand is continuous, but presents a set of 'caverns' of various sizes each of which is connected to several others by narrower channels. The 'caverns' or pores have a certain size distribution which may approximate to the normal or may be of Poisson type. They are also randomly distributed in space. The theoretical problem which presents itself is the computation of the permeability of such a random array, and it appears to have received but scant attention from mathematical statisticians. We have therefore perforce proceeded by a bold application of simple concepts. In so far as the results have been satisfactory,

it may perhaps be held that our concepts have merely led to a method of computing permeability, the real justification of which is empirical.

Let us take $f(r)$ as the distribution function which describes the pore space of an uncemented sand, i.e. $f(r) \delta r$ is the fraction of the total apparent volume which is occupied by pores of 'radius' range r to $r + \delta r$. A cross-section of this sand will exhibit a fraction of the total area, also given by $f(r) \delta r$, devoted to this pore group. Let us now consider a sand column of unit cross-sectional area. A fracture at any chosen plane normal to the length will exhibit two similar faces showing similar pore-size distributions; the continuous column may be regarded as the random juxtaposition of these two faces. If a_ρ is the area devoted to pores of radius ρ to $\rho + \delta r$ and a_σ the area devoted to pores in the range σ to $\sigma + \delta r$, then

$$a_\rho = f(\rho) \delta r, \quad a_\sigma = f(\sigma) \delta r.$$

Hence, on remaking the continuous column, the area of cross-section devoted to the pore sequence $\rho \rightarrow \sigma$ is given by

$$a_{\rho \rightarrow \sigma} = f(\rho) \delta r f(\sigma) \delta r. \quad (10)$$

We now make two assumptions, which introduce errors of opposite sign. The first is that all the effective resistance to flow in the sequence is confined to the smaller of the pores, say σ , owing to the operation of the factor r^4 in Poisseuille's equation. This assumption results in an over-estimation of the contribution of this sequence to the total permeability, but frees us from the necessity of considering the probability of more than one of the smaller pores leading into the same larger pore. The second assumption is that the only contribution to permeability is by a direct sequence of the kind described, i.e. we ignore by-passing sequences of, maybe, several pores. This results in an under-estimation tending to compensate for the previous over-estimation. The distribution of pore sequences given by (10) is repeated wherever the cross-section is taken, and is characteristic of the material. The number of pores of size σ accommodated in the area $a_{\rho \rightarrow \sigma}$ is proportional to σ^{-2} , the rate of flow in each is, if we apply Poisseuille's equation, proportional to σ^4 per unit potential gradient, whence the contribution δK which this group of sequences makes to the total permeability K is given by

$$\delta K = \sigma^2 M f(\rho) \delta r f(\sigma) \delta r, \quad (11)$$

and the total permeability due to all the possible sequences by

$$K = M \sum_{\rho=0}^{\rho=R} \sum_{\sigma=0}^{\sigma=R} \sigma^2 f(\rho) \delta r f(\sigma) \delta r. \quad (12)$$

In (12) the factor σ^2 must be replaced by ρ^2 in any term of the summation in which $\rho < \sigma$. For any given moisture content the summation is stopped at that pore size, R , appropriate to the largest pore which remains full of water. The constant M can hardly be calculated; it will in fact be determined in this paper by matching theoretical and experimental curves at a single point.

For a non-shrinking material such as sand, the distribution function $f(r)$ may be derived satisfactorily from the moisture characteristic, i.e. from the curve relating moisture content to the hydrostatic pressure at that content (Childs 1940, 1942).

The amount by which this pressure is less than atmospheric is a measure of the curvature of the air-water interface in equilibrium in the pore space, and therefore a measure of the largest cells which have not yet been emptied of water. The volume of water removed by a given increment of suction is thus a measure of the pore space occupied by pores of a known range of sizes.

Where there are very many groups of pores, the calculation indicated by (12) may clearly be a laborious task, and we therefore present a form of tabulation which much reduces the work and which also clearly indicates its course. In the first place, a pore which is emptied at pressure deficiency P will have a 'radius' which we can relate to $2S/P$, since an air-water interface of that radius is involved. We therefore replot the moisture characteristic in the form *moisture content* against $1/P$. There is no need to invoke the surface tension S , since (12) in any case includes the constant M which is obtained by empirical matching, so that for brevity we may refer to $1/P$ as the 'radius'. From the resulting curve we tabulate the pore volume to be assigned to each size group, i.e. $f(r) dr$.

Let the pore groups have mean radii, beginning with the smallest, of $a, b, c, \dots$ respectively, contributing porosity elements $\alpha, \beta, \gamma, \dots$ (i.e. α is written for $f(a) dr$ and so on). Then (12) becomes

$$K = M(a^2\alpha^2 + b^2\beta^2 + c^2\gamma^2 + \dots + 2a^2\alpha\beta + 2a^2\alpha\gamma + \dots + 2b^2\beta\gamma + \dots), \quad (13)$$

the series terminating at the largest water-occupied pore size. It should be noted that the ' r^2 ' factor, i.e. a^2, b^2 , etc., in each term is always the radius of the smaller pore in the sequence. A table is constructed, as shown in table 1. Columns 1 and 2 are abstracted from the transformed moisture characteristic, adopting a suitable pore-size interval dr for the groups. The rest of the table is self-explanatory; it need only be pointed out that column 4 is obtained by summing column 3 down to and including the row being filled in, and column 6 is obtained from column 5 similarly, the result in this case being doubled. Column 8 is obtained by summing column 7. Column 9, multiplied by M , gives the expression (13) appropriate to the limiting pore radius indicated in column 1, i.e. appropriate to a moisture content which may be regarded as the sum of column 2 down to the row indicated but which is naturally most simply read off from the moisture characteristic. As an example, table 2 gives

TABLE 1

1	2	3	4	5
r	$f(r) \delta r$	$r^2 f(r) \delta r$	$\Sigma r^2 f(r) \delta r$	$f(r) \delta r \Sigma r^2 f(r) \delta r$
a	α	$a^2 \alpha$	$a^2 \alpha$	$a^2 \alpha^2$
b	β	$b^2 \beta$	$a^2 \alpha + b^2 \beta$	$a^2 \alpha \beta + b^2 \beta^2$
c	γ	$c^2 \gamma$	$a^2 \alpha + b^2 \beta + c^2 \gamma$	$a^2 \alpha \gamma + b^2 \beta \gamma + c^2 \gamma^2$
etc.	etc.	etc.	etc.	etc.
	6	7	8	9
	$2\Sigma\{f(r) \delta r \Sigma r^2 f(r) \delta r\}$	$r^2\{f(r) \delta r\}^2$	$\Sigma r^2\{f(r) \delta r\}^2$	$(= 6 - 8)$
	$2a^2 \alpha^2$	$a^2 \alpha^2$	$a^2 \alpha^2$	K/M
	$2(a^2 \alpha^2 + b^2 \beta^2 + a^2 \alpha \beta)$	$b^2 \beta^2$	$a^2 \alpha^2 + b^2 \beta^2$	$a^2 \alpha^2 + b^2 \beta^2 + 2a^2 \alpha \beta$
	$2(a^2 \alpha^2 + b^2 \beta^2 + c^2 \gamma^2$	$c^2 \gamma^2$	$a^2 \alpha^2 + b^2 \beta^2 + c^2 \gamma^2$	$a^2 \alpha^2 + b^2 \beta^2 + c^2 \gamma^2$
	$+ a^2 \alpha \beta + a^2 \alpha \gamma + b^2 \beta \gamma)$			$+ 2a^2 \alpha \beta + 2a^2 \alpha \gamma + 2b^2 \beta \gamma$
	etc.	etc.	etc.	etc.

TABLE 2. THE MOISTURE CHARACTERISTIC OF 1 TO 1/2 MM. SAND AND THE DERIVATION OF THE PORE-SIZE DISTRIBUTION

data interpolated from the graph of moisture content against reciprocal of tension

experimental results			data interpolated from the graph of moisture content against reciprocal of tension			
moisture content (ml./100 ml.)	tension (cm. of water)	reciprocal of tension	reciprocal of tension	moisture content (% volume)	volume contribution to pore group $\int f(r)dr$	mean radius of pore group r
35.67	0.0	∞	0.02625	3.42	0.00	0.0000
35.59	2.4	0.4167	0.02875	3.44	0.02	0.0275
35.26	6.5	0.1539	0.03125	3.48	0.04	0.0300
35.10	10.4	0.09615	0.03375	3.53	0.05	0.0325
34.75	14.3	0.06993	0.03625	3.65	0.08	0.0350
31.41	16.3	0.06135	0.03875	4.00	0.35	0.0375
27.21	17.1	0.05848	0.04125	4.61	0.61	0.0400
22.32	17.7	0.05650	0.04375	5.58	0.97	0.0425
19.38	18.05	0.05541	0.04625	7.00	1.42	0.0450
14.15	18.65	0.05362	0.04875	8.72	1.72	0.0475
10.27	19.8	0.0505	0.05125	11.14	2.42	0.0500
6.675	21.8	0.04587	0.05375	15.30	4.16	0.0525
4.385	24.7	0.04049	0.05625	22.00	6.70	0.0550
3.59	28.4	0.03521	0.05875	27.95	5.95	0.0575
3.50	32.5	0.03077	0.06125	31.37	3.42	0.0600
3.43	36.4	0.02747	0.06375	34.35	2.02	0.0625
			0.06625	34.60	0.25	0.0650
			0.06875	34.71	0.11	0.0675
			0.07125	34.80	0.09	0.0700
			0.0775	34.89	0.09	0.0744
			0.0825	34.97	0.08	0.0800
			0.0875	35.02	0.05	0.0850
			0.0925	35.07	0.05	0.0900
			0.1050	35.15	0.08	0.0988
			0.1150	35.20	0.05	0.1100
			0.1250	35.23	0.03	0.1200
			0.1350	35.25	0.02	0.1300
			0.1650	35.30	0.05	0.1500
			0.1950	35.35	0.05	0.1800
			0.2250	35.39	0.04	0.2100
			0.2550	35.43	0.04	0.2400
			0.3250	35.50	0.07	0.2900
			0.3950	35.57	0.07	0.3600
			0.4650	35.61	0.04	0.4300
			0.5350	35.64	0.03	0.5000
			0.6050	35.66	0.02	0.5700
			0.6750	35.67	0.01	0.6400

PORE-SIZE DISTRIBUTION OF 1 TO $\frac{1}{2}$ MM. SAND

1	2	3	4	5	6	7	8	9	10
									$K(\text{c.g.s.})$
0.0000	0.00	0.000,0153	0.000,0153	0.000,000,305	0.000,000,610	0.000,000,305	0.000,000,305	0.000,000,305	1.69×10^{-11}
0.0275	0.02	360	513	2,05	4,71	1,44	1,74	2,97	1.65×10^{-10}
0.0300	0.04	528	1041	5,21	15,12	2,64	4,38	10,74	5.97
0.0325	0.05	980	2021	16,17	47,46	7,84	12,22	35,34	1.96×10^{-9}
0.0350	0.08								
0.0375	0.35	0.000,492	0.000,694	0.000,243	0.000,533	0.000,172	0.000,184	0.000,349	1.94×10^{-8}
0.0400	0.61	976	1,670	1,019	2,571	595	779	1,792	9.95
0.0425	0.97	1,752	3,422	3,32	9,21	1,695	2,474	6,74	3.74×10^{-7}
0.0450	1.42	2,878	6,300	8,95	27,11	4,09	6,56	20,55	1.14×10^{-6}
0.0475	1.72	3,88	10,18	17,52	62,15	6,68	13,24	48,91	2.72
0.0500	2.42	0.006,05	0.016,23	0.039,26	0.140,7	0.014,65	0.027,89	0.112,8	6.26×10^{-6}
0.0525	4.16	11,47	27,70	0.115,2	0.371,1	47,70	75,59	0.295,5	1.64×10^{-5}
0.0550	6.70	20,28	47,98	0.321,5	1.014	0.136,0	0.211,6	0.802	4.45
0.0575	5.95	19,69	67,67	0.402,8	1.820	0.117,2	0.328,8	1.491	8.28
0.0600	3.42	12,32	79,99	0.273,7	2.367	0.042,2	0.371,0	1.996	1.11×10^{-4}
0.0625	2.02	0.007,89	0.087,88	0.177,6	2.722	0.015,9	0.386,9	2.335	1.296×10^{-4}
0.0650	0.25	1.06	88,94	22,2	2.767	3	0.387,2	2.380	1.322
0.0675	0.11	51	89,45	9.8	2.786	1	0.387,3	2.399	1.332
0.0700	0.09	44	89,89	8,1	2.803	0	0.387,3	2.416	1.342
0.0744	0.09	50	90,39	8,1	2.819	0	0.387,3	2.432	1.351
0.0800	0.08	0.000,51	0.090,90	0.007,3	2.833	0.000,0	0.387,4	2.446	1.359×10^{-4}
0.0850	0.05	36	91,26	4,6	2.843	0	0.387,4	2.456	1.364
0.0900	0.05	41	91,67	4,6	2.852	0	0.387,4	2.465	1.369
0.0988	0.08	78	92,45	7,4	2.866	1	0.387,5	2.478	1.376
0.1100	0.05	61	93,06	4,7	2.875	0	0.387,5	2.487	1.382
0.1200	0.03	0.000,43	0.093,49	0.002,8	2.880	0.000,0	0.387,5	2.492	1.384×10^{-4}
0.1300	0.02	34	93,83	1,9	2.883	0	0.387,5	2.495	1.386
0.1500	0.05	1.13	94,96	4,7	2.893	1	0.387,6	2.505	1.391
0.1800	0.05	1.62	96,58	4,8	2.902	1	0.387,7	2.514	1.396
0.2100	0.04	1.76	98,34	3,9	2.911	1	0.387,8	2.523	1.401
0.2400	0.04	0.002,30	0.100,64	0.004,0	2.919	0.000,1	0.387,9	2.531	1.405×10^{-4}
0.2900	0.07	5.89	0.106,53	7,5	2.934	4	0.388,3	2.546	1.414
0.3600	0.07	9.08	0.115,61	8,1	2.950	6	0.388,9	2.561	1.423
0.4300	0.04	7,40	0.123,01	4,9	2.960	3	0.389,2	2.571	1.428
0.5000	0.03	7,50	0.130,51	3,9	2.969	2	0.389,4	2.580	1.433
0.5700	0.02	0.006,50	0.137,01	0.002,7	2.974	0.000,1	0.389,5	2.585	$*1.435 \times 10^{-4}$
0.6400	0.01	4,10	0.141,11	1,4	2.977	0	0.389,5	2.587	1.436
∞	0.00	0	0.141,11	0	2.977	0	0.389,5	2.587	1.436

The columns are numbered to agree with those of table 1.

* M chosen to match observed permeability at this point.

the moisture characteristic of the coarsest material shown in figure 1 and its conversion into a pore-size distribution, whilst table 3 shows the calculation of the appropriate permeability against moisture content curve illustrated in figure 2.

THE MEASUREMENT OF PERMEABILITY OF UNSATURATED SAND

The difficulties inherent in the measurement of the permeability of unsaturated porous materials are readily seen. The necessary potential gradient generally involves a pressure gradient, and a gradient of pressure deficiencies implies, *ipso facto*, a gradient of moisture content. A flow column can hardly be sampled without spoiling it for subsequent measurements, and so one must resort to indirect measurements, such as by measurement of the apparent electric capacitance of a condenser of which the porous material forms the dielectric. The presence of a moisture gradient then complicates matters, and we have ourselves tried intricate systems of guard rings to isolate a restricted region of approximately uniform moisture content. However, the ultimate solution of the problem was very simple, although limited to structureless materials such as sands and dusts. It has been shown (Childs 1945) that when water flows down a sufficiently long soil column to a water table, the moisture content is sensibly uniform over an appreciable length of the column, the zones of variable moisture content being limited at the lower end to the neighbourhood of the water table in a way which depends upon the pore-size distribution and at the upper end to a zone in which is localized any intermittency of water supply. In the zone of uniform moisture content there is no pressure contribution to the potential gradient, which is therefore purely gravitational and exactly known. The moisture content adjusts itself to provide the necessary permeability to conduct the imposed flow with the gravitational gradient of potential. To calculate the permeability we need to measure only the rate of flow and the area of cross-section of the column, whilst the uniform moisture content lends itself to estimation by indirect electrical methods using a simple calibrated cell.

The flow column used was a glass tube some 3 m. in length, of internal diameter about 2 cm., and containing a removable section approximately 16 cm. in length at about 65 cm. from the upper end. Two electrodes of aluminium foil, of dimensions 3 by 3.4 cm., were sealed to the outside of and half-way along this section, forming a split tube; these electrodes formed the plates of a condenser, the dielectric of which was in part the glass walls of the tube and in part the moist sand or dust. The condenser formed part of the tuned circuit of a triode oscillator tuned to about 1 Mcycle. A small variable condenser was connected in parallel with this cell and with a larger condenser. A second triode oscillator was arranged to beat with the first. Changes of apparent capacitance of the cell resulted in changes of beat frequency, and were measured by bringing the first circuit back to resonance by adjusting the small parallel condenser. Readings on this condenser were interpreted in terms of moisture content of the dielectric of the cell by an initial direct calibration, the cell being removed from the main flow column for determination of moisture content by direct weighing.

The lower end of the tube was immersed below a water surface, and flow to the upper end was controlled by glass syphons from a constant-head reservoir. The tube

was filled in stages with sand or dust by settlement through shallow water; in this way both trapped air and fractionation of the sample were avoided. The rate of flow was then adjusted in steps to secure the required states of moisture content, starting with saturation and proceeding to decreased contents. It was found that, using ordinary tap water, it was impossible to maintain the initial rate of flow in saturated sands. Constancy of flow at the initial maximum rate was possible when boiled distilled water was used, so it was probable that the trouble was due to dissolved air coming out of solution in the sand pores and thereby reducing the moisture content and permeability, a conclusion which was supported by the moisture content 'meter'. It was not practicable to use boiled distilled water for all saturated columns, but tap water was satisfactory once it was known that the true permeability was that at 'zero time'. The trouble did not arise with unsaturated sands, presumably because the continuous air-filled pore space provided a leak path for the escape of momentarily trapped air soon after its emergence from solution.

The whole flow tube and its ancillary equipment was mounted on a beam suspended at its upper end from a ceiling pivot. In this way inclinations from the vertical of up to 60° could be secured, thus providing potential gradients adjustable at will down to $g/2$. By this means one could test for the validity of Darcy's law in the prevailing conditions; if the law is obeyed, then the rate of flow is proportional to the potential gradient, other factors remaining constant.

EXPERIMENTAL RESULTS

The collected results are presented in figures 1 to 3. The moisture characteristics of the materials used, from which the permeabilities were computed, are shown in figure 1. These materials include a sand fraction separated between round-holed sieves of 1 and 0.5 mm. apertures, a second sand fraction separated between the

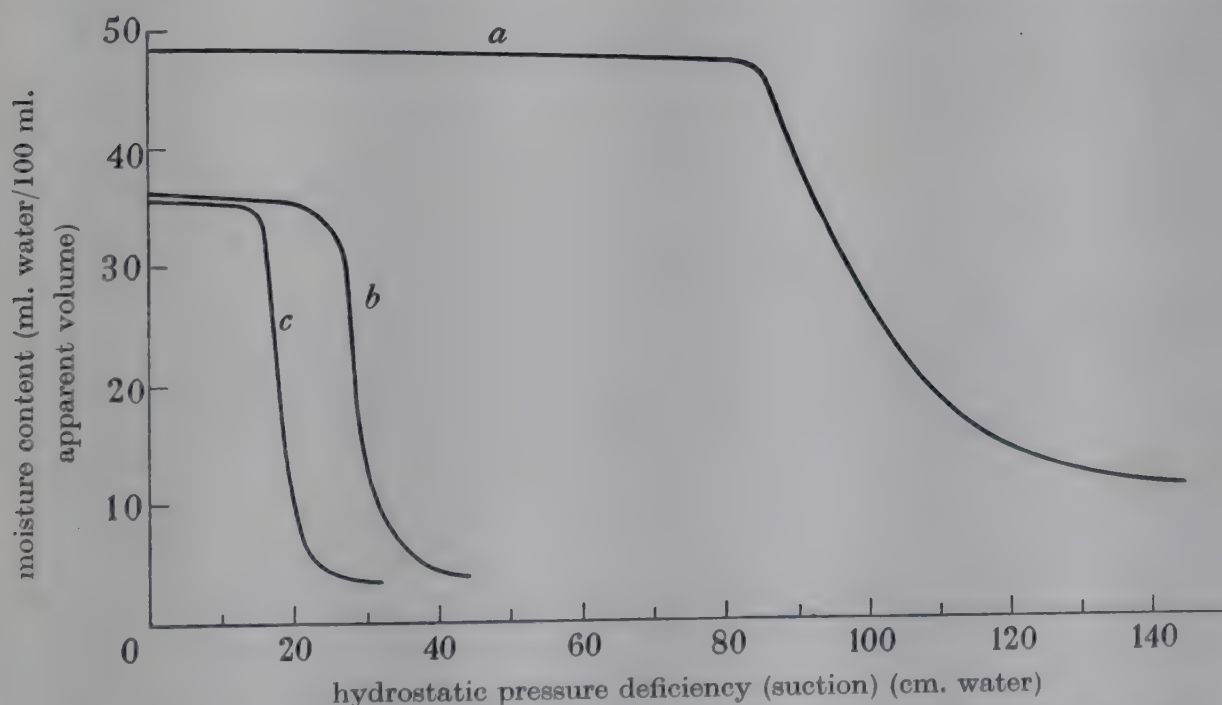


FIGURE 1. Moisture characteristics: *a*, slate dust; *b*, $\frac{1}{2}$ to $\frac{1}{4}$ mm. sand fraction; *c*, 1 to $\frac{1}{2}$ mm. sand fraction.

finer of these two sieves and one with 0.25 mm. square holes (wire-woven), and a slate dust passing a sieve with 0.125 mm. square holes but containing no particles smaller than 0.04 mm., as calculated from the velocity of sedimentation. This information serves only to describe the materials roughly, not to specify them; it is our argument that the specification is provided by the moisture characteristics.

In figure 2 are shown the observed and calculated permeabilities of these three materials, as functions of moisture content. The constant M in equation (12), required in order to calculate the permeability in absolute terms, is obtained for all materials by matching the observed and calculated curves for the coarsest material at the point indicated in figure 2. This set of curves includes also, for comparison, a computed curve for the coarsest material based on Kozeny's equation. This will be discussed further in due course, but it will be obvious that the agreement with observation is not close.

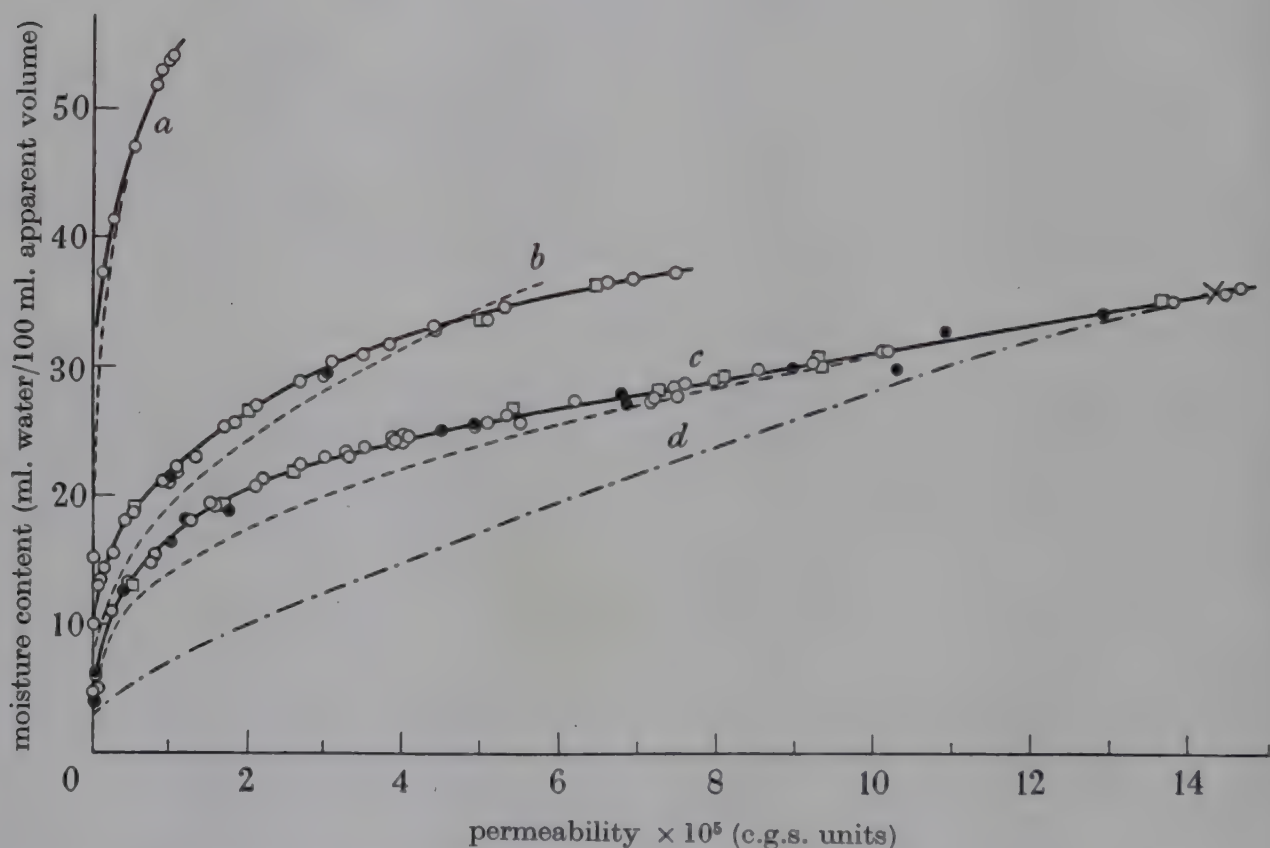


FIGURE 2. Observed and calculated permeabilities: *a*, slate dust; *b*, $\frac{1}{2}$ to $\frac{1}{4}$ mm. sand fraction; *c*, 1 to $\frac{1}{2}$ mm. sand fraction; *d*, 1 to $\frac{1}{2}$ mm. sand fraction. Experimental curve ———. Computed curve: Childs & George - - - -; after Kozeny — — —. Experimental points potential gradient: g $\circ$; $\frac{3}{4}g$ $\bullet$; $\frac{1}{2}g$ $\square$. Experimental and computed curves matched at $\times$.

DISCUSSION

We have referred already to the imperfect nature of the agreement between permeabilities calculated from Kozeny's equation (3) and those observed. The difficulty of making the comparison at all is another ground for doubting the validity of the equation. The term (V/A) in equation (3) was calculated for each stage from the moisture characteristic, it being taken as axiomatic that empty

pores were as ineffective conductors of water as if they had been filled with solids, i.e. V includes solids and empty pores. Referring both V and A to unit apparent volume of sand, we may substitute $(1 - p)$ for V , whence equation (3) becomes

$$K = Dp^3/A^2, \quad (14)$$

where A is now the area of the solid-liquid interface per unit apparent volume and p is the effective (water-filled) porosity. This area A is derived by summing the contributions of each pore-size group and involves a knowledge of the contributions of those groups which lie outside the range of the moisture characteristic, contributions which, owing to small pore size, are far from negligible, even though the contributions to porosity are minute. The trend of the permeability against moisture content curve at the low moisture content end and within the range of the moisture characteristic depends very much upon the way in which one approximates in these circumstances. In deriving the curve shown in figure 2 we have assumed, from the fact that the observed permeability is negligible at a certain lower limit of moisture content, that this limit is the effective datum for the estimate of porosity. It is the necessity for such an assumption which we refer to as indicating the inadequacy of the Kozeny type formula. No such assumption is required for the calculation of permeability from (12), the group contributions becoming sensibly zero within the range of the moisture characteristics.

The limitations of our conception will be obvious. We have confined ourselves to a cross-sectional factor and have ignored the effect of the different lengths of cells of different sizes. A sequence of two small pores, for example, occupies a much smaller proportion of the total volume than does a sequence of two larger pores, and should presumably be weighted less on that account. This may well account for the relative over-estimate of permeability at low moisture contents, where the small pores become significant. It might seem that if one were to consider sequences of very many pores instead of pore pairs, the frequency of occurrence would be concentrated about an average sequence, departures from this being rare, so that weighting of different sequences would not arise. Other difficulties, however, have prevented much progress on these lines; it is clearly very artificial to ignore all but the smallest pore in a sequence of many, and the problem of series-parallel pore connexions arises in an acute form. However, as a matter of empirical trial, it is found that a better agreement between observed and calculated permeability against moisture-content curves is obtained with our method of computation for pore sequences up to four. The error then tends to reverse in sign, the calculated permeabilities at low moisture content being relatively too low.

We have been concerned until now with permeability, i.e. with the velocity of flow per unit potential gradient. One is very often more conscious of moisture movement as a result of a gradient of moisture content, as, for example, when irrigation water redistributes itself down the soil profile during the first few days after the irrigation. The problem is then one of diffusion with a coefficient of diffusion which varies with both time and depth. We may write Darcy's law for one dimension (vertical flow)

$$dQ/dt = -Kd\phi/dz, \quad (15)$$

where dQ/dt is the rate of flow (ml./sec.) in the vertical direction z . The hydraulic potential is the sum of gravitational and 'capillary' components, and may be written

$$\phi = P + gs z,$$

whence

$$d\phi/dz = dP/dz + gs. \quad (16)$$

Here P is the pressure component (a suction in unsaturated materials) and s is the density of the water. The moisture profile represents the moisture content, m , as a function of z , whilst, from the moisture characteristic, it is also expressible as a function of P . Hence, combining (14) and (15), we may write

$$\begin{aligned} dQ/dt &= -K(dP/dm \cdot dm/dz + gs) \\ &= -\{\kappa dm/dz + Kgs\}. \end{aligned} \quad (17)$$

The first term, indicating a rate of movement of moisture down a concentration gradient, is clearly a diffusion term with a coefficient of diffusion, κ , which can be calculated, for a given moisture content, from the moisture characteristic, since this latter permits us to calculate both factors in κ , namely the permeability K and the slope of the characteristic dP/dm . In figure 3 are shown the diffusion coefficients of the coarser sand and the slate dust, plotted as functions of moisture content. It will often arise, in heavier soils with low K and high dP/dm (and always with horizontal movement), that the second term of (16) will be negligible compared with the first; in such a case there would appear to be a prospect of forecasting the movement, for example, of irrigation water from measurements of the moisture characteristics of profile samples, always bearing in mind that the problem may also be

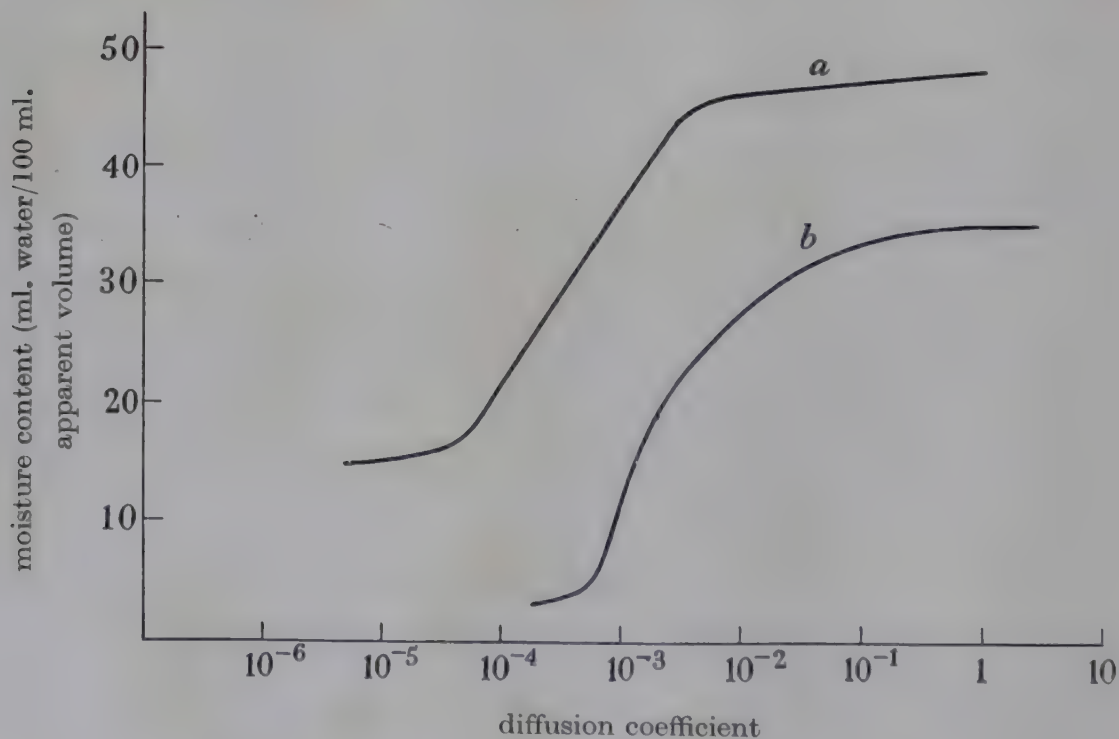


FIGURE 3. Moisture content of *a*, slate dust, and *b*, 1 to $\frac{1}{2}$ mm. sand fraction plotted against the diffusion coefficients.

complicated by hysteresis in the characteristic. Such a forecast requires the solution of the non-linear diffusion problem, and attention is now being turned in that direction.

REFERENCES

- Carman, P. C. 1938 *J. Soc. Chem. Ind., Lond.*, **57**, T 225.
 Carman, P. C. 1939 *J. Soc. Chem. Ind., Lond.*, **58**, T 1.
 Childs, E. C. 1940 *Soil Sci.* **50**, 239.
 Childs, E. C. 1942 *Soil Sci.* **53**, 79.
 Childs, E. C. 1945 *Soil Sci.* **59**, 405.
 Christiansen, J. E. 1944 *Soil Sci.* **58**, 355.
 Donat, J. 1937 *Trans. 6th Comm. Int. Soc. Soil Sci. B*, p. 423.
 Fair, G. M. & Hatch, L. P. 1933 *J. Amer. Wat. Wks Ass.* **25**, 1551.
 Fancher, G. H., Lewis, J. A. & Barnes, K. B. 1933 *Bull. Pa. St. Coll.* no. 12.
 Feng, C. L. & Browning, G. M. 1946 *Proc. Soil Sci. Soc. Amer.* **11**, 67.
 Kozeny, J. 1927 *S.B. Akad. Wiss. Wien, A*, **136**, 271.
 Lea, F. M. & Nurse, R. W. 1939 *J. Soc. Chem. Ind., Lond.*, **58**, T 277.
 Muskat, M. 1937 *The flow of homogeneous fluids through porous media*, p. 70. New York: McGraw-Hill Book Company, Inc.
 Rigden, P. J. 1943 *J. Soc. Chem. Ind., Lond.*, **62**, T 1.
 Rigden, P. J. 1947 *J. Soc. Chem. Ind., Lond.*, **66**, T 130.
 Slichter, C. S. 1897-8 *U.S. Geol. Survey, 19th Ann. Rep.* pt. 2, p. 295.
 Terzaghi, C. 1925 *Engng News Rec.* p. 832.
 Zunker, F. 1933 *Trans. 6th Comm. Int. Soc. Soil Sci. B*, p. 18.

On the spontaneous magnetic field in a conducting liquid in turbulent motion

By G. K. BATCHELOR, *Trinity College, University of Cambridge*

(Communicated by Sir Geoffrey Taylor, F.R.S.—Received 24 November 1949)

Several recent investigations in geophysics and astrophysics have involved a consideration of the hydrodynamics of a fluid which is a good electrical conductor. In this paper one of the problems which seem likely to arise in such investigations is discussed. The fluid is assumed to be incompressible and in homogeneous turbulent motion, and externally imposed electric and magnetic fields are assumed to be absent. The equations governing the interaction of the electromagnetic field and the turbulent motion are set up with the same assumptions as are used to obtain the Maxwell and current flow equations for a metallic conductor. It is shown that the equation for the magnetic field is identical in form with that for the vorticity in a non-conducting fluid; immediate deductions are that lines of magnetic force move with the fluid when the conductivity is infinite, and that the small-scale components of the turbulence have the more powerful effect on the magnetic field.

The first question considered is the stability of a purely hydrodynamical system to small disturbing magnetic fields, and it is shown that the magnetic energy of the disturbance will increase provided the conductivity is greater than a critical value determined by the viscosity of the fluid. The rate of growth of magnetic energy is approximately exponential, with a doubling time which can be simply related to the properties of the turbulence. General mechanical considerations suggest that a steady state is reached when the magnetic field has as much energy as is contained in the small-scale components of the turbulence. Estimates of this amount of energy and of the region of the spectrum in which it will lie are given in terms of observable properties of the turbulence.

1. INTRODUCTION

The general subject of the hydrodynamics of fluids which are electrically conducting is now of considerable interest in view of its relevance to various problems arising in geophysics and astrophysics. As examples of the latter, it is sufficient to mention the work of Bullard (1949) on the origin of the earth's magnetic field, Alfvén's theory (Alfvén 1943) of sunspots, the work of Walén (1946) on the general magnetic field of the sun, and finally the theories of the origin of cosmic rays put forward by Fermi (1949) and by Alfvén (1949). Speculation about the explanation of these different physical phenomena would be less difficult and more profitable if there existed a background of general analysis concerning electromagnetic hydrodynamics, comparable with that which already exists for ordinary hydrodynamics. A good beginning to a systematic study of the subject was made by Hartmann (1937) and Hartmann & Lazarus (1937), but this work does not seem to have been continued.

This paper is intended to make a small contribution to the systematic study of the general problem, without reference to particular physical circumstances. The only assumption about the physical conditions is that the usual Maxwell equations and constitutive relations are valid, apart from modifications required by the hydrodynamic motion. In this sense, the subject of the paper is the electromagnetic hydrodynamics of metal-type conductors. The particular intention is to consider the interaction of the electromagnetic and hydrodynamic forces in the absence of any imposed electric or magnetic field or distribution of electric charge and in the absence of appreciable electromagnetic radiation. The hydrodynamic motion will be assumed to be turbulent (as is usually the case under natural conditions) and homogeneous, i.e. the average properties of the motion are independent of position. The assumed conditions correspond to the physical problem of a body of conducting fluid which has initially been set into turbulent motion and has had time to distribute the turbulent energy uniformly (on the average) over space. Actually it will be found that it is the small-scale components of the turbulence which are of most interest in connexion with the electromagnetic interaction, and for these small-scale components the assumption of homogeneity will be a good approximation for all kinds of turbulent motion.

The author's interest in the problem was aroused at the symposium on 'Problems of motion of gaseous masses of cosmical dimensions' arranged jointly by the International Unions for Astronomy and for Applied Mechanics, in Paris in August 1949. This paper is a development of a contribution made at the symposium and owes much to the stimulating discussions in which the author was privileged to take part.

2. THE GOVERNING ELECTROMAGNETIC AND HYDRODYNAMIC EQUATIONS

Since the rates of change of the electric and magnetic field vectors $\mathbf{E}$ and $\mathbf{H}$ are determined by the instantaneous distribution and motion of negative charges relative to positive charges, irrespective of how that distribution and motion are produced, Maxwell's equations are formally unaffected by the hydrodynamic motion. Hence,

on the assumption that the dielectric constant K and the magnetic permeability μ have the same values at all points of the fluid,

$$K\nabla \cdot \mathbf{E} = 4\pi\rho_e, \quad (2.1)$$

$$\nabla \cdot \mathbf{H} = 0, \quad (2.2)$$

$$\frac{K}{c} \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - 4\pi\mathbf{j}, \quad (2.3)$$

$$\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E}, \quad (2.4)$$

where ρ_e and $\mathbf{j}$ are the excess charge and current densities; Gaussian units have been used, with $\mathbf{j}$ (and, later, the conductivity σ) in e.m.u. The current flow equation must be modified, however, since there is now a convection current $\rho_e \frac{\mathbf{v}}{c}$, where $\mathbf{v}$ is the local mass velocity of the fluid, and since the electric field is now not the only force acting on charges at rest with respect to the fluid. Between collisions with molecules of the fluid, one e.m.u. of charge is accelerated initially by the force $c\mathbf{E} + \mu \mathbf{v} \times \mathbf{H}$, so that provided the time between collisions is small compared with the Larmor rotation time,* the conduction current is $\sigma(c\mathbf{E} + \mu \mathbf{v} \times \mathbf{H})$ and the current flow equation becomes

$$\mathbf{j} = \sigma(c\mathbf{E} + \mu \mathbf{v} \times \mathbf{H}) + \rho_e \frac{\mathbf{v}}{c}. \quad (2.5)$$

The resultant equation in $\mathbf{H}$ alone is found to be

$$\frac{\partial \mathbf{H}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{H}) = \frac{1}{4\pi\mu\sigma} \left(\nabla^2 \mathbf{H} - \frac{\mu K}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} + \frac{4\pi}{c} \nabla \times \rho_e \mathbf{v} \right). \quad (2.6)$$

If $\mathbf{v} = 0$, (2.6) reduces to the familiar equation of wireless telegraphy. If there are no boundaries producing high-frequency electromagnetic oscillations it will be generally true that

$$\frac{K}{4\pi\sigma c^2} \left| \frac{\partial^2 \mathbf{H}}{\partial t^2} \right| \ll \left| \frac{\partial \mathbf{H}}{\partial t} \right|,$$

in which case (2.6) reduces (again with $\mathbf{v} = 0$) to a heat conduction type of equation, and in this form is used, for example, to analyze skin effect in conductors carrying alternating current. Sources of high-frequency electromagnetic oscillations, if present at all, can usually be ignored in the kind of hydrodynamic problem under consideration because of the small amounts of energy likely to be involved, so that the second term on the right side of (2.6) will be neglected hereafter. The importance of the last term on the right side can be determined from the equation for the rate of change of charge contained in a moving element of fluid, obtained from equations (2.1) to (2.5), viz.

$$\frac{\partial \rho_e}{\partial t} + \nabla \cdot (\rho_e \mathbf{v}) = -\frac{4\pi\sigma c^2}{K} \rho_e - c\sigma\mu \nabla \cdot (\mathbf{v} \times \mathbf{H}).$$

* This condition for the usual analysis to be applicable may not be satisfied when the density of the fluid is very small, as in the clouds of interstellar gas, and when the magnetic field has appreciable energy.

The term $-\frac{4\pi\sigma c^2}{K}\rho_e$ represents an exponential rate of decay of ρ_e (with a time of relaxation which is extremely small for the conducting materials likely to be of interest in the present connexion), whereas $-\sigma\mu\nabla\cdot(\mathbf{v}\times\mathbf{H})$ is not directly related to ρ_e . Hence, if it is imagined that the motion has existed for a time long enough for the initial distribution of ρ_e to be irrelevant, the average order of magnitude of ρ_e is comparable with that of $\frac{K\mu}{4\pi c}\nabla\cdot(\mathbf{v}\times\mathbf{H})$. The term $\frac{4\pi}{c}\nabla\times(\rho_e\mathbf{v})$ inside the bracket in (2.6) is therefore of the order of $K\mu\nabla\times\left[\frac{\mathbf{v}}{c}\nabla\cdot\left(\frac{\mathbf{v}}{c}\times\mathbf{H}\right)\right]$. Since this is less than $|\nabla^2\mathbf{H}|$ by a factor of order $|\mathbf{v}|^2/c^2$, it follows that the third term on the right side of (2.6) can be neglected. Accordingly the equation to be studied herein is

$$\frac{\partial\mathbf{H}}{\partial t}-\nabla\times(\mathbf{v}\times\mathbf{H})=\lambda\nabla^2\mathbf{H}\quad\left(\lambda=\frac{1}{4\pi\mu\sigma}\right). \quad (2.7)$$

Turning now to a consideration of the bulk motion of the fluid, it will be assumed that the fluid is incompressible, i.e. that the density ρ is constant, and so

$$\nabla\cdot\mathbf{v}=0. \quad (2.8)$$

Admittedly, this will be a poor approximation in cases where the turbulent velocities are of the order of the velocity of sound, as in interstellar gas clouds, and (2.8) must then be regarded as a first step towards a more complete analysis. The equation of momentum consists of the ordinary Navier-Stokes equation modified to take account of the electromagnetic body forces, viz. the electrostatic force and the Lorentz deflecting force,

$$\frac{\partial\mathbf{v}}{\partial t}+(\mathbf{v}\cdot\nabla)\mathbf{v}=\frac{1}{\rho}(\rho_e\mathbf{E}+\mu\mathbf{j}\times\mathbf{H})-\frac{1}{\rho}\nabla p+\nu\nabla^2\mathbf{v}, \quad (2.9)$$

where ν is the kinematic viscosity and p is the hydrodynamic pressure.

The electromagnetic force per unit volume may be written as

$$\begin{aligned} \rho_e\mathbf{E}+\mu\mathbf{j}\times\mathbf{H} &= \frac{K}{4\pi}\mathbf{E}(\nabla\cdot\mathbf{E})+\frac{\mu K}{4\pi c}\frac{\partial\mathbf{H}\times\mathbf{E}}{\partial t}-\frac{\mu}{4\pi}\mathbf{H}\times(\nabla\times\mathbf{H})-\frac{K}{4\pi}\mathbf{E}\times(\nabla\times\mathbf{E}) \\ &= \frac{\mu K}{4\pi c}\frac{\partial\mathbf{H}\times\mathbf{E}}{\partial t}+\frac{K}{4\pi}\nabla[(\cdot\mathbf{E})\mathbf{E}-\tfrac{1}{2}|\mathbf{E}|^2]+\frac{\mu}{4\pi}\nabla[(\cdot\mathbf{H})\mathbf{H}-\tfrac{1}{2}|\mathbf{H}|^2]. \end{aligned} \quad (2.10)$$

Thus only the first term on the right side, which represents the rate of decrease of linear momentum of the electromagnetic field, is a true body force. The remaining terms represent the force resulting from a system of stresses, the stress dyadic being

$$\frac{K}{4\pi}\mathbf{E}\mathbf{E}+\frac{\mu}{4\pi}\mathbf{H}\mathbf{H}-\frac{1}{8\pi}(K|\mathbf{E}|^2+\mu|\mathbf{H}|^2)\mathbf{I}, \quad (2.11)$$

where $\mathbf{I}$ is the unit diagonal dyadic. This stress system can be regarded as being produced by the tubes of electric and of magnetic force exerting longitudinal tensions

$\frac{K}{8\pi}|\mathbf{E}|^2$ and $\frac{\mu}{8\pi}|\mathbf{H}|^2$ per unit area and lateral pressures of the same magnitudes,

respectively. In general, regions where the curvature of the lines of electric and of magnetic force are high will be regions where the stress system is changing rapidly, i.e. where the electromagnetic forces on the fluid are large.

3. ANALOGY BETWEEN MAGNETIC FIELD AND VORTICITY

Consider for the moment a non-conducting, incompressible fluid. It may be shown from the equations of continuity and momentum that the vector vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{v}$ satisfies the equations

$$\nabla \cdot \boldsymbol{\omega} = 0, \quad (3.1)$$

$$\frac{\partial \boldsymbol{\omega}}{\partial t} - \nabla \times (\mathbf{v} \times \boldsymbol{\omega}) = \nu \nabla^2 \boldsymbol{\omega}. \quad (3.2)$$

These two equations are identical in form with equations (2.2) and (2.7), the viscosity ν replacing the constant λ which has the character of a magnetic diffusivity. There is thus a formal analogy between the two solenoidal vectors $\boldsymbol{\omega}$ and $\mathbf{H}$, provided $\boldsymbol{\omega}$ refers to the motion of a non-conducting liquid and $\mathbf{H}$ to the motion of a conducting liquid.

Many of the results concerning vorticity in classical hydrodynamics can now be interpreted in terms of magnetic field in the electromagnetic hydrodynamic problem. For instance, Helmholtz's result that lines of vorticity move with the fluid when the fluid is frictionless shows that the lines of magnetic force move with the fluid when $\lambda = 0$, i.e. when the conductivity is infinite. In this case the lines of magnetic force are embedded in the medium; any tendency for them to move relative to the medium is immediately checked by very large eddy currents. An understanding of the mechanical actions represented by the terms in equation (2.7) can be obtained from previous studies of the corresponding terms in the vorticity equation (3.2). The term $\lambda \nabla^2 \mathbf{H}$ clearly represents a diffusion or spreading of the magnetic field. The other term contributing to $\partial \mathbf{H} / \partial t$ can be written

$$\nabla \times (\mathbf{v} \times \mathbf{H}) = (\mathbf{H} \cdot \nabla) \mathbf{v} - (\mathbf{v} \cdot \nabla) \mathbf{H}. \quad (3.3)$$

$(\mathbf{v} \cdot \nabla) \mathbf{H}$ represents a convection of $\mathbf{H}$, and may be added to $\partial \mathbf{H} / \partial t$ to form the so-called time differentiation following the motion of the fluid. $(\mathbf{H} \cdot \nabla) \mathbf{v}$ represents a rate of change of $\mathbf{H}$ which is due to relative motion of neighbouring points on a line of magnetic force. This motion in general consists of a rotation, which leaves $|\mathbf{H}|$ unchanged, and an extension, which alters the value of $|\mathbf{H}|$. When $\mathbf{v}$ is so distributed as to increase the length of the line of magnetic force, $|\mathbf{H}|$ increases, and vice versa.

It is now possible to perceive the general features of the interplay of hydrodynamic and electromagnetic effects. The medium grips the lines of magnetic force, with a firmness measured by the magnitude of σ , and moves them about; in so doing it may either extend or contract the lines and thereby increases or decreases the energy of the magnetic field. The magnetic lines of force are in tension and tend to contract, thereby accelerating the fluid. Other manifestations of the system of electromagnetic stresses likewise tend to convey kinetic energy to the fluid. In a study of turbulent motion, the task is to obtain information about the statistical flow of energy between different parts of the whole dynamical system.

4. THE CRITERION FOR GROWTH OF ELECTROMAGNETIC ENERGY IN A TURBULENT MOTION

Since the problem of interest herein concerns turbulent motion of a conducting fluid in the absence of any imposed electric or magnetic field, the electromagnetic energy, if it is to be present at all, must arise spontaneously from the hydrodynamic motion. It can be supposed that in reality there will always be present small disturbing electric and magnetic fields produced by factors extraneous to the fluid motion, so that it is necessary to determine whether the average effect of the turbulence is to amplify or suppress these small disturbing fields. An answer to this question can be given if the disturbing fields are assumed to be (self-consistent) random functions of position, with probability distributions the same at all points. In other words both the turbulent velocity and the electromagnetic field vectors are assumed to be stationary random functions. Of course the disturbing electric and magnetic fields will not usually have this character immediately after they are introduced. However, the turbulent motion will distort them and redistribute them in a random fashion so that ultimately they will approximate to stationary random functions; whether they have suffered decay or amplification in becoming stationary random functions is not relevant to their eventual fate.

The equation for the rate of change of the average amount of magnetic energy in unit volume of fluid is found from (2.7) to be

$$\frac{1}{2} \frac{d \overline{|\mathbf{H}|^2}}{dt} = \overline{\mathbf{H} \cdot \nabla \times (\mathbf{v} \times \mathbf{H})} + \lambda \overline{\mathbf{H} \cdot \nabla^2 \mathbf{H}}. \quad (4.1)$$

With the use of (3.3) and (2.8) and the assumption that $\mathbf{H}$ is a stationary random function, this becomes

$$\begin{aligned} \frac{1}{2} \frac{d \overline{|\mathbf{H}|^2}}{dt} &= \overline{\mathbf{H} \cdot [(\mathbf{H} \cdot \nabla) \mathbf{v}]} - \overline{(\mathbf{v} \cdot \nabla) |\mathbf{H}|^2} - \lambda \overline{|\nabla H_i|^2} \\ &= \overline{|\mathbf{H}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_H} \right)_H} - \lambda \overline{|\nabla H_i|^2}, \end{aligned} \quad (4.2)$$

where the suffix H means a component in the direction of the line of magnetic force and the repeated suffix i indicates summation over three orthogonal components. Thus $\overline{|\mathbf{H}|^2}$ will increase or decrease according to the relative importance of the two terms in (4.2). The second term on the right side is always negative and describes the average rate of conversion of magnetic energy into Joule heat. The first term represents the flow of energy to or from the turbulent motion, and the form of the term shows that this flow depends on the extension and contraction of the lines of magnetic force produced by the turbulence. All terms in (4.2) are of the same order of smallness in $|\mathbf{H}|$, so that $\overline{|\mathbf{H}|^2}$ will continue to grow or to decay until such time as there is a change in the statistical characteristics of $\mathbf{v}$ or $\mathbf{H}$.

Again the analogy with the vorticity is useful. In the absence of electromagnetic effects, the mean square vorticity of a homogeneous turbulent flow satisfies the equation

$$\frac{1}{2} \frac{d \overline{|\boldsymbol{\omega}|^2}}{dt} = \overline{|\boldsymbol{\omega}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_\omega} \right)_\omega} - \nu \overline{|\nabla \boldsymbol{\omega}_i|^2}, \quad (4.3)$$

and the same mechanical effects will operate as in the equation (4.2). Now it is well known in the subject of turbulence that on the average particles of the fluid tend to diffuse apart and that this process lengthens lines which move with the fluid. Lines of vorticity do not move with the fluid exactly when $\nu \neq 0$, but inasmuch as they do so approximately, they tend to lengthen and it is always found experimentally (for air) that

$$\overline{|\omega|^2 \left(\frac{\partial \mathbf{v}}{\partial x_\omega} \right)_\omega} > 0.$$

Stretching the lines of vorticity leads to a positive contribution to $\frac{1}{2} \frac{d \overline{|\omega|^2}}{dt}$ for the same reasons (formally) that stretching the lines of magnetic force increases the magnetic energy, both angular momentum and magnetic flux being conserved by the motion of the fluid. It is also found experimentally (Batchelor & Townsend 1947) that when the Reynolds number of the turbulence is sufficiently large,* the two terms on the right side of (4.3) are each large compared with their difference. In other words the processes of decay of vorticity by viscous spreading and the amplification of vorticity by extension of the vortex lines are approximately in equilibrium.

It will be noticed that the contribution of the electromagnetic forces to the equation of hydrodynamic momentum (2.9) is of the second degree in $|\mathbf{E}|$ and $|\mathbf{H}|$; consequently, while the electromagnetic energy of the disturbance is still small, the turbulence is approximately unaffected by the fact that the fluid is conducting. The result stated above that the contributions to the rate of change of mean-square vorticity are in equilibrium is therefore applicable to the turbulence in the conducting fluid unless and until the electromagnetic energy becomes large. The mechanical processes acting on the vorticity are identical with those acting on the magnetic field, except that the diffusivity coefficients have different values. Hence, provided the statistical characteristics of the distribution of $\mathbf{H}$ are approximately

the same (apart from absolute value) as those of ω the two contributions to $\frac{d \overline{|\mathbf{H}|^2}}{dt}$ will be in equilibrium if $\lambda = \nu$; if $\lambda > \nu$, the damping of the magnetic energy will be stronger than the amplification by hydrodynamic stretching of the lines of magnetic force and $\overline{|\mathbf{H}|^2}$ will decay to zero; and if $\lambda < \nu$, the inflow of energy from the turbulence will be greater than the Joule heat loss and $\overline{|\mathbf{H}|^2}$ will increase.

Concerning the above proviso, it will not be true in general that the statistical characteristics of $\mathbf{H}$ and ω are similar immediately after the disturbing magnetic field is introduced, but a state of approximate similarity will soon be set up. The process of stretching of magnetic lines not only makes a positive contribution to $\frac{d \overline{|\mathbf{H}|^2}}{dt}$, it also transfers magnetic energy from large-scale variations of $\mathbf{H}$ to small-

* Extrapolation of the wind-tunnel data suggests that $(\overline{|\mathbf{v}|^2})^{\frac{1}{2}} L / \nu$ should be greater than about 10^5 , where L is a measure of the size of the large components of the turbulence containing the bulk of the energy $\frac{1}{2} \overline{|\mathbf{v}|^2}$. This criterion will almost certainly be satisfied strongly in naturally occurring turbulence owing to its large scale. For instance, it would be satisfied in atmospheric turbulence if the energy-containing motions were on a scale of 10 m. with a root-mean-square value of $|\mathbf{v}|$ of 0.15 m./sec.

scale variations of $\mathbf{H}$. The transfer of energy to smaller scales is limited, however, by the non-existence of the stretching process for very small scales. In this way the bulk of the magnetic energy finds its way into the wave numbers containing most of the mean-square vorticity, and the spatial distributions of $\mathbf{H}$ and $\boldsymbol{\omega}$ will become approximately similar; from this stage onward the above criterion is applicable.

Thus the criterion for the energy of the disturbing magnetic field to increase is that $\lambda < \nu$, i.e. that

$$4\pi\mu\sigma\nu > 1. \quad (4.4)$$

As an example, one finds for mercury at normal temperatures and pressures that $4\pi\mu\sigma\nu \approx 1.5 \times 10^{-7}$ ($\sigma \approx 10^{-5} \text{ cm.}^{-2}\text{sec.}$, $\nu \approx 1.2 \times 10^{-3} \text{ cm.}^2\text{sec.}^{-1}$), so it is quite certain that magnetic energy will not increase spontaneously in mercury in turbulent motion. The criterion (4.4) is clearly very restrictive and will be satisfied only in very extreme circumstances. σ and ν increase with increase of the mean free paths of the carriers of charge and mass respectively, so that a rarified ionized gas probably has the best chance of satisfying (4.4). There is then the difficulty that the assumed condition, that the time between collisions of the carriers of charge is small compared with the Larmor rotation time, may not be satisfied after the magnetic energy has been amplified beyond a certain value. At least it will be true that (4.4) is the criterion for increase of $|\overline{\mathbf{H}}|^2$, while $|\overline{\mathbf{H}}|^2$ is smaller than this critical value.

5. THE RATE OF GROWTH OF THE MAGNETIC ENERGY

Assume in this and the following section that the fluid is such as to satisfy the criterion (4.4). This implies that the conductivity is large, with the usual consequences. For instance, whatever the initial conditions, $\mathbf{E} \cdot \mathbf{H}$ quickly becomes zero everywhere. The current density $\mathbf{j}$ is finite, so that (2.5) shows that $|\mathbf{E}|$ is of the order of $\mu \left| \frac{\mathbf{v}}{c} \right| |\mathbf{H}|$. Since $|\mathbf{v}| \ll c$, it follows that there is much more energy in the magnetic field than in the electric field. For the same reason the electromagnetic force (2.10) reduces to the term in $\mathbf{H}$ alone, viz. $-\frac{\mu}{4\pi} \mathbf{H} \times (\nabla \times \mathbf{H})$. The problem thus reduces to the interaction of the fields of $\mathbf{v}$ and $\mathbf{H}$.

Consider now the rate at which $|\overline{\mathbf{H}}|^2$ will increase in such a fluid. Equation (4.2) may be written in the form

$$\frac{1}{2} \frac{d \log |\overline{\mathbf{H}}|^2}{dt} = \frac{|\overline{\mathbf{H}}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_H} \right)_H}{|\overline{\mathbf{H}}|^2} - \lambda \frac{|\overline{\nabla H_i}|^2}{|\overline{\mathbf{H}}|^2}. \quad (5.1)$$

Provided the statistical characteristics of $\mathbf{H}$, apart from its magnitude, do not change appreciably with time—and there is no apparent reason why they should, once the effect of the initial form of the disturbing magnetic field has died out—the right side of (5.1) is constant. Thus $|\overline{\mathbf{H}}|^2$ increases exponentially with time. A good estimate of the constant value of (5.1) can be obtained by dropping the second term on the right side, since this will be important only in the uncommon special case in which λ is close to ν . The remaining term depends on the stretching characteristics of the

turbulent motion. These stretching characteristics affect the vorticity and the magnetic field in the same way, so that

$$\frac{\overline{|\mathbf{H}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_H} \right)_H}}{\overline{|\mathbf{H}|^2}} \approx \frac{\overline{|\boldsymbol{\omega}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_\omega} \right)_\omega}}{\overline{|\boldsymbol{\omega}|^2}}; \quad (5.2)$$

the equality is exact if the difference between the statistical characteristics of $\mathbf{H}$ and $\boldsymbol{\omega}$ is one of magnitude alone. Experimentally one finds (Batchelor & Townsend 1947) that at high Reynolds numbers the right side of (5.2) is proportional to $(\epsilon/\nu)^{\frac{1}{2}}$ for all states of the turbulence,* where ϵ is the rate of dissipation of kinetic energy by viscosity per unit mass of the fluid; also the constant of proportionality is found to be of order unity. Thus the solution of (5.1) is

$$\overline{|\mathbf{H}|^2} = \overline{|\mathbf{H}|_0^2} \exp \left[\alpha \left(\frac{\epsilon}{\nu} \right)^{\frac{1}{2}} t \right], \quad (5.3)$$

where $\overline{|\mathbf{H}|_0^2}$ is the value of $\overline{|\mathbf{H}|^2}$ at the chosen origin of t , and α is a constant of order unity.

It may be useful, for the practical application of (5.3), to remark that an order of magnitude estimate of ϵ may be obtained from the formula

$$\epsilon \propto (\overline{|\mathbf{v}|^2})^{\frac{3}{2}}/L, \quad (5.4)$$

which is generally accepted in turbulence research on both experimental and theoretical grounds (for the definition of L , see the footnote to § 4).

The rate of increase of magnetic energy given by (5.3) cannot persist indefinitely, of course. The ultimate steady value of $\overline{|\mathbf{H}|^2}$ is discussed in the following section.

6. THE EQUILIBRIUM BETWEEN THE MAGNETIC AND HYDRODYNAMIC SYSTEMS

It has been remarked that the stretching process by which the magnetic energy is amplified is similar to that which keeps the mean-square vorticity at a steady value against the damping effect of viscosity. Those parts of a tube of magnetic force which contain larger amounts of energy will also be those which have been more extended. Now the small-scale components of the turbulence are more effective in stretching lines than the large-scale components; or more precisely, the part of the spectrum of turbulence which is responsible for most of the extension of the magnetic lines is not the part containing the energy of the turbulence, but is the part containing most of the vorticity. At high Reynolds numbers, these two ranges of the spectrum, which can conveniently be termed small and large wave-numbers, respectively, will be fairly separate and non-overlapping (Taylor 1938), so that the distinction is important. Thus a spectral analysis of the spatial variation of $\mathbf{H}$ would reveal that the stretching process was feeding energy into those wave-numbers which also contain most of the vorticity of the turbulence.

The implication is that the large wave-number components of the turbulence are wholly responsible for the building up of the magnetic field. This building up process

* As predicted by Kolmogoroff's hypotheses about turbulent motion; see § 6.

ceases when the magnetic stress is sufficiently large to pass as much energy to the turbulent motion as the magnetic field is receiving. The details of this latter process will be described by the momentum equation (2.9) and (2.10), viz.

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = \frac{\mu}{4\pi\rho} (\mathbf{H} \cdot \nabla) \mathbf{H} - \frac{1}{\rho} \nabla \left(p + \frac{\mu |\mathbf{H}|^2}{8\pi} \right) + \nu \nabla^2 \mathbf{v}. \quad (6.1)$$

The following hypothesis, although not proven rigorously, therefore seems plausible: *The ultimate balance between the magnetic and hydrodynamic systems is such that the large wave-number components contain comparable amounts of kinetic and magnetic energy.*

When the equilibrium is reached, the lines of magnetic force have sufficient tension to resist the turbulent stretching mechanism, which, as has been pointed out, has available the kinetic energy associated with the large wave numbers. Since the vorticity—and other quantities associated with the large wave numbers—is approximately in statistical equilibrium when the Reynolds number is large, irrespective of the rate of change of total turbulent energy, so too is the magnetic field. Hence the following equations will apply when the magnetic field has reached equilibrium:

$$\frac{1}{2} \frac{d}{dt} \overline{|\mathbf{H}|^2} = \overline{|\mathbf{H}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_H} \right)_H} - \lambda \overline{|\nabla H_i|^2} = 0, \quad (6.2)$$

$$\frac{1}{2} \frac{d}{dt} \overline{|\boldsymbol{\omega}|^2} = \overline{|\boldsymbol{\omega}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_\omega} \right)_\omega} - \nu \overline{|\nabla \omega_i|^2} = 0, \quad (6.3)$$

$$\begin{aligned} \frac{1}{2} \frac{d}{dt} \overline{|\mathbf{v}|^2} &= -\frac{\mu}{4\pi\rho} \overline{\mathbf{v} \cdot [\mathbf{H} \times (\nabla \times \mathbf{H})]} + \nu \overline{\mathbf{v} \cdot \nabla^2 \mathbf{v}} \\ &= -\frac{\mu}{4\pi\rho} \overline{|\mathbf{H}|^2 \left(\frac{\partial \mathbf{v}}{\partial x_H} \right)_H} - \nu \overline{|\nabla v_i|^2}, \\ \text{and on using (6.2),} \quad &= -\frac{\mu\lambda}{4\pi\rho} \overline{|\nabla H_i|^2} - \nu \overline{|\nabla v_i|^2}. \end{aligned} \quad (6.4)$$

According to the above hypothesis $\frac{\mu}{4\pi} \overline{|\nabla H_i|^2}$ and $\rho \overline{|\nabla v_i|^2}$ are of the same order of magnitude, so that the total amount of energy lost to heat is not changed in order of magnitude by the presence of the magnetic field. If there is a steady source of turbulent energy, a steady state for all components of the system will be set up and the total dissipation to heat given by (6.4) will be equal to the rate of gain of energy from the source. The energy-spectrum flow diagram will be roughly as shown in figure 1.

To estimate the value of the magnetic energy in the steady state, use may be made of Kolmogoroff's postulates about the large wave-number components of turbulence at high Reynolds number (Kolmogoroff 1941). These postulates, which have been substantially confirmed by experiment, state that the various statistical quantities associated strongly with the large wave-number components of the turbulence (including the mean rate of extension of lines moving with the fluid and the rate of viscous dissipation of kinetic energy, ϵ) are determined by the parameters

ϵ and ν alone. In particular, the amount of energy associated with the large wave-number components is of order $(\epsilon\nu)^{\frac{1}{2}}$ per unit mass of fluid. This will still be true of a conducting fluid when the magnetic field has reached a steady statistical state, and hence on the hypothesis made earlier in this section

$$\frac{\mu}{8\pi} |\overline{\mathbf{H}}|^2 \sim \rho(\epsilon\nu)^{\frac{1}{2}} \quad (6.5)$$

in the steady state. Likewise the length representing both the large wave-number components of the turbulence and those of the magnetic field is of order $(\nu^3/\epsilon)^{\frac{1}{2}}$ so that

$$\frac{\mu}{8\pi} |\overline{\nabla \times \mathbf{H}}|^2 \sim \rho \frac{\epsilon}{\nu}, = \rho |\overline{\boldsymbol{\omega}}|^2. \quad (6.6)^*$$

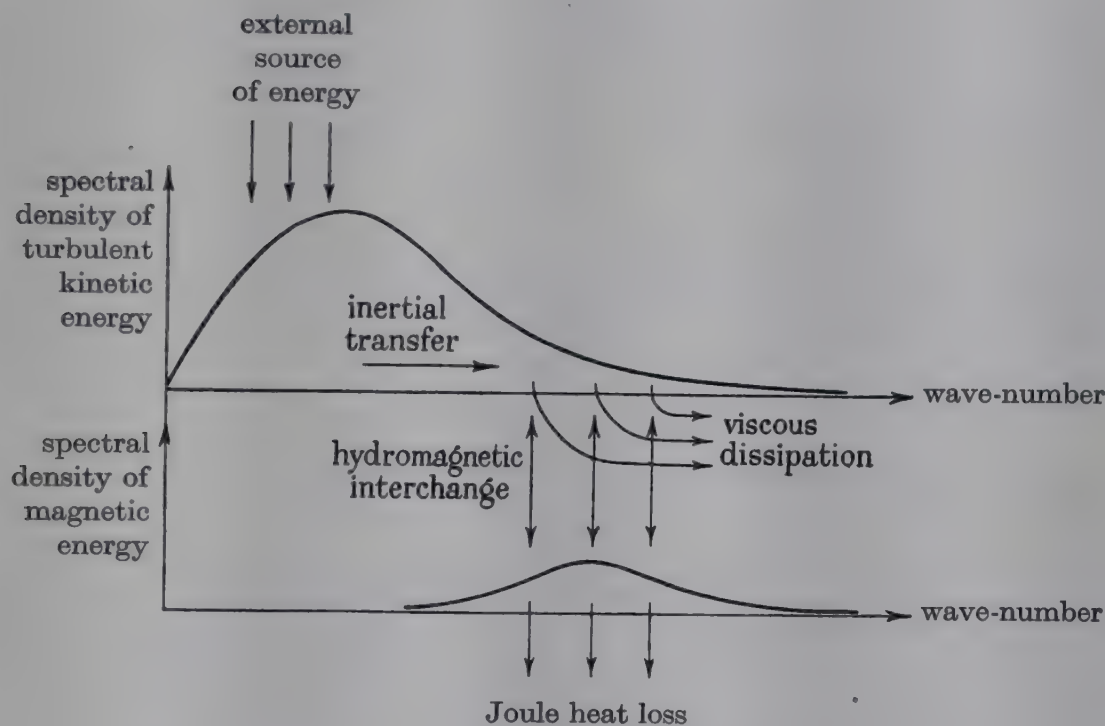


FIGURE 1. Diagram of the flow of energy in the steady state.

The magnetic energy will increase at a rate which decreases as it approaches the asymptotic value (6.5), but a rough estimate of the time to reach the steady state

from an initial magnetic energy $\frac{\mu}{8\pi} |\overline{\mathbf{H}}|_0^2$ is given by (5.3) as

$$t \sim \left(\frac{\nu}{\epsilon}\right)^{\frac{1}{2}} \log \frac{8\pi\rho(\epsilon\nu)^{\frac{1}{2}}}{\mu |\overline{\mathbf{H}}|_0^2}. \quad (6.7)$$

The meaning of the hypothesis made in this section can be otherwise stated as follows. The local instantaneous stretching or deformation tensor (of which a typical component is $\partial v_i/\partial x_j$) is approximately constant over a region of the fluid which has linear dimensions of order l , say. The velocity differences over this region are of order $l \left| \frac{\partial v_i}{\partial x_j} \right|$, so that the average amount of kinetic energy available for balancing

* It is an exact result for homogeneous turbulence that $\epsilon = \nu |\overline{\boldsymbol{\omega}}|^2$.

the stored potential energy of the stretched strings which represent the magnetic field is of order $l \left(\frac{\partial v_i}{\partial x_j} \right)^2$. Since l is effectively a measure of the diameter of eddies which make the greatest contribution to $\left(\frac{\partial v_i}{\partial x_j} \right)^2$,

$$l^2 \left(\frac{\partial v_i}{\partial x_j} \right)^2 = \text{kinetic energy in the large wave-number components} \\ \sim (\epsilon \nu)^{\frac{1}{2}}, \quad \text{from Kolmogoroff's postulates,}$$

and this magnitude thus describes the magnetic energy in the steady state. It is true that the total kinetic energy of the turbulence may be much greater than $(\epsilon \nu)^{\frac{1}{2}}$, but this energy is not available for interchange with the magnetic energy owing to the fact that it is associated chiefly with small wave numbers. The greater part of the turbulent kinetic energy serves only to translate the lines of magnetic force and $|\mathbf{v}^2|$ is an irrelevant parameter for the magnetic field (except inasmuch as it plays a part in determining $|\boldsymbol{\omega}|^2$). Alfvén (1949) and Fermi (1949) have come to the contrary conclusion that $|\mathbf{H}|^2$ is comparable with $|\mathbf{v}|^2$ in the steady state, for reasons which appear to be based on the conception of equipartition of energy; however, this conception in its simple form is scarcely applicable to the above dissipative system.

7. SUMMARY OF CONCLUSIONS

A small random magnetic field introduced into a conducting liquid (for which Ohm's law holds) in homogeneous turbulent motion will be amplified if $4\pi\mu\sigma\nu > 1$. The initial rate of growth of the magnetic energy will be exponential, with a doubling time of order $(\nu/\epsilon)^{\frac{1}{2}}$. It is suggested that ultimately the magnetic field reaches a statistically steady state with energy of order $\rho(\epsilon\nu)^{\frac{1}{2}}$ per unit volume of fluid. This magnetic energy will have a spectral distribution which is concentrated in the neighbourhood of wave numbers of order $(\epsilon/\nu^3)^{\frac{1}{2}}$.

In a practical case, the relation (5.4) will usually enable an estimate of the order of magnitude of ϵ to be made from fairly crude non-statistical observational data.

An extension of this analysis which would increase greatly its range of applicability is to the case of a (non-metal-type) fluid in which the time between collisions of the carriers of charge is not small compared with the Larmor rotation time for magnetic fields of the strength estimated herein.

REFERENCES

- Alfvén, H. 1943 *Ark. Mat. Astr. Fys.* **29**B, no. 2 and **29**A, no. 12.
 Alfvén, H. 1949 *Phys. Rev.* **75**, 1732.
 Batchelor, G. K. & Townsend, A. A. 1947 *Proc. Roy. Soc. A*, **190**, 534.
 Bullard, E. C. 1949 *Proc. Roy. Soc. A*, **197**, 433.
 Fermi, E. 1949 *Phys. Rev.* **75**, 1169.
 Hartmann, J. 1937 *Math.-fys. Medd.* **15**, no. 6.
 Hartmann, J. & Lazarus, F. 1937 *Math.-fys. Medd.* **15**, no. 7.
 Kolmogoroff, A. N. 1941 *C.R. Acad. Sci. U.R.S.S.* **30**, 301 and **32**, 16.
 Taylor, G. I. 1938 *Proc. Roy. Soc. A*, **164**, 15.
 Walén, C. 1946 *Ark. Mat. Astr. Fys.* **33**A, no. 18.

World-structure and non-Euclidean honeycombs

By H. S. M. COXETER,* *University of Toronto*
AND G. J. WHITROW, *Imperial College of Science, London*

(Communicated by E. A. Milne, F.R.S.—Received 26 November 1949)

Milne (1934) described a one-dimensional system of discrete particles in uniform relative motion such that the aspect of the whole system is the same from each particle. The purpose of the present paper is to construct analogous systems in two and three dimensions. If the uniformly moving observers regraduate their clocks so as to describe each other as relatively stationary, the private Euclidean spaces of the Special Theory of Relativity become public hyperbolic space. This point of view leads to a discussion of uniform honeycombs in hyperbolic space, four of which were discovered by Schlegel (1883, p. 444). One of the new honeycombs, called $\{4, 4, 3\}$, has for its vertices the points whose four co-ordinates are proportional to the integral solutions of the Diophantine equation

$$t^2 - x^2 - y^2 - z^2 = 1.$$

As a by-product, a simple set of generators and generating relations are obtained for the group of all integral Lorentz transformations (Schild 1949, p. 39). Another by-product is the enumeration of those groups generated by reflexions in hyperbolic space whose fundamental regions are tetrahedra of finite volume. The work culminates in the discovery of a point-distribution whose mesh is seven times as close as that of $\{4, 4, 3\}$, though apparently still far too coarse to be of direct cosmological significance. It follows that some irregularity in the distribution of the extragalactic nebulae is almost certainly *geometrically* inevitable.

1. INTRODUCTION

Most of the models which have been investigated in recent theoretical studies of world-structure are homogeneous. Comparatively little progress has been made in the study of non-homogeneous models, e.g. models in which matter is concentrated in discrete 'nebulae'. McCrea (1931) investigated a 'cubical' universe in which matter was regarded as concentrated at the sixteen vertices of a four-dimensional cube. In the Einstein universe the spatial section is the surface of the four-dimensional sphere

$$x^2 + y^2 + z^2 + w^2 = a^2,$$

and McCrea considered the case where this surface is replaced by

$$x^{2n} + y^{2n} + z^{2n} + w^{2n} = a^{2n},$$

where n is a large positive integer. As n increases this approximates more and more closely to a hypercube. Since the curvature is small except in the neighbourhood of the sixteen points $(\pm a, \pm a, \pm a, \pm a)$, it follows from general relativity that the matter necessary to produce this curvature must be mainly concentrated at these points. As McCrea remarked, 'the interest of this particular model is that it reproduces to some extent an essential feature of actual space in which the matter is more or less symmetrically concentrated into isolated nebulae or galaxies'.

Tolman (1934) showed that under certain simplifying assumptions a completely homogeneous world-model is unstable under perturbations in density, so that any

* Elected F.R.S., 16 March 1950.

local tendency to expand would result in further expansion and any local tendency to contract in further contraction. Consequently, a homogeneous model will tend to become non-homogeneous. Recently, Tolman's researches have been extended by Omer (1949), who has developed a non-homogeneous cosmological model as a higher approximation to the solution of the cosmological problem than that given by the homogeneous models. He has assumed, however, a spherically symmetrical distribution of matter about a central observer, and remarks, 'The assumption of a central observer may seem to be philosophically objectionable. However, with the present observational data there is no other logical position in which to locate him.'

An alternative approach to the problem of obtaining a higher approximation than the homogeneous 'smoothed-out' world-models was initiated by Milne (1934). This method is akin to McCrea's in so far as there is no appeal to the assumption of a central observer, the model being 'regular' in the sense in which a cube or any of the Platonic polyhedra are regular, but the matter in it is isolated into discrete point concentrations. At first Milne confined his attention to the preliminary problem of constructing a one-dimensional universe of discrete particles in uniform relative motion such that the distribution is described in the same way by equivalent observers attached to each constituent particle. The generalization to the case in which the equivalent particles are in any prescribed form of relative motion was further examined by Whitrow (1935, p. 259). The results obtained by Milne and Whitrow, however, are not exhaustive. They correspond to systems of collinear points evenly spaced, and other solutions are admissible corresponding to systems of collinear points spaced alternately at two different distances, e.g. like the corners of a rectangle on its circumcircle.

Milne (1935, p. 280) announced that 'it appears impossible to construct an abstract system of discretely separated particles satisfying the cosmological principle in three dimensions', i.e. such that each describes the system, from his own point of view, in the same way. It was assumed, of course, that the individual particles were in uniform relative motion, in accordance with the simplest interpretation of the observed red-shifts in the spectra of the nebulae and their correlation with assigned distances. If, however, the particles had been assumed to be relatively stationary, then the problem would have been equivalent to the purely geometrical problem of constructing a homogeneous honeycomb in the particular space considered, e.g. Euclidean space. In this space it is well known that such honeycombs exist, e.g. the ordinary cubic lattice. More generally we could have the transforms of an arbitrary point under any one of the 230 space-groups of crystallography.

Whitrow (1935, p. 256) showed that different motion-patterns ('equivalences') could be transformed into each other by appropriate regraduation of time-scale, and Milne (1937) drew attention to the logarithmic transformation relating a uniformly expanding equivalence and a static one. In particular, he found that if the same convention for determining distances in terms of clock-readings and signal velocity (the 'radar convention') were retained, the Euclidean space which he had adopted for his uniformly expanding world-model would be transformed into a hyperbolic space of constant curvature for the static description of the same model. The natural unit of distance in this space is equal to the distance travelled by light

since the Creation, i.e. between 9^9 and 10^{10} light-years, if the age of the universe is that number of years. Whitrow (1939) showed that *part* of this investigation of Milne was identical with the investigations of Varićak (1910) and Sommerfeld (1909). Although they had not anticipated the idea of transforming the time-scale, they had shown that the Lorentz formulae were compatible with a public kinematic (velocity) space of Lobatschewskian type. It then became evident that the problem of constructing in Euclidean (private) space a system of discrete particles in uniform relative motion such that observers attached to each particle would describe the system in the same way, from his own point of view, could be reduced to the purely geometrical problem of constructing a *homogeneous honeycomb in hyperbolic space*. It is well known that this problem is soluble (Schlegel 1883). Consequently it was apparent that Milne had been mistaken in his announcement, and that there must have been a flaw in his alleged proof (Milne 1935, pp. 357–360).

The flaw occurred in the following argument. Suppose A , B and C to be any three particles in uniform relative motion. ' A observes B and C to be moving with certain velocities in his experience. Hence C sees B to be moving, in C 's experience, with a velocity that can be calculated by transforming B 's velocity from A 's description of it to C 's. Since the descriptions of the whole system from A , B and C are to be identical, A must see a particle to be moving with the same velocity in A 's experience as C sees B to be moving with in C 's experience.' In developing this idea Milne assumed that A and C assign the same speed and *orientation* to these two particles respectively, and deduced that, for this to be possible, A , B and C must be collinear. The assumption about *orientation* is, however, unnecessary and, strictly speaking, meaningless, for there is no 'absolute background' to the system. Only by invoking a background can we maintain that the various aspects of a cube, say, from its respective vertices are not identical. If, however, we consider only the mutual relations of the vertices then the aspect of them all is the same from each. Similarly, in Milne's problem all we need, and indeed can, consider are the mutual relations of the particles as seen from each.* Milne's proof is, therefore, invalid, and systems of discrete uniformly moving particles obeying the 'cosmological principle' are theoretically possible.

A problem which is closely related to ours has been studied recently by Schild (1949). He has investigated integral Lorentz transformations in discrete *space-time*. He considered all events in Minkowskian space-time whose four co-ordinates (t, x, y, z) are integers (the velocity of light being taken as unity). These events form

* These considerations apply equally to the construction of continuous world-models: Milne's method of applying the cosmological principle to construct the substratum and the associated statistical systems is open to the same criticism concerning orientation (Milne 1935). Relaxing his implicit assumption of an absolute background for his co-ordinates (x, y, z) , we suggest that it is preferable to replace his cosmological principle by an effectively equivalent combination of the sample principle (Whitrow 1936, 1937), the principle of spherical symmetry about each fundamental observer (Walker 1937) and criteria for the geometry of light-paths similar to those of Helmholtz-Lie (Robertson 1935). This method not only leads to some mathematical simplification of Milne's procedure but is *epistemologically superior*, as it is no longer necessary to presuppose any geometrical background other than that provided by the world-model itself.

a hypercubic lattice in space-time. His object, however, was different from ours, for he remarks: 'It seems likely that a physical theory based on a discrete space-time background will be free of the infinities which trouble contemporary quantum mechanics.' Schild is concerned with lattices in Minkowskian space-time, and we are concerned with discrete points in the associated velocity-space. In Schild's case there is a least non-zero interval ϵ between lattice-points, which is determined physically, not geometrically. He remarks: 'In any physical theory based on our model, ϵ would probably be of the general order of magnitude of the classical electron radius (approximately 10^{-13} cm.).'

The model discussed here, of course, is intended for comparison with the observable universe of nebulae, and if we identify the discrete particles occurring in it with the major nebulae (e.g. the Galaxy, M 31, etc.), then the least spatial interval between non-coincident points should be of the order of 10^6 light-years. It is well known, however, that the extragalactic nebulae display a marked clustering tendency, and if we identify our particles with clusters the minimum spatial interval may be taken to be of the order of (at most) 10^7 light-years. In comparing our model with observation the significant feature is the ratio of the distance between neighbouring lattice-points and the radius of curvature of the hyperbolic space. In Milne's theory, as in many other relativistic cosmologies, this radius is of the order of a few thousand million light-years. Thus the ratio of the minimum distance to the radius should be of the order of one part in a few thousands, or alternatively of one part in a few hundreds. As we shall see, these ratios are considerably smaller than those pertaining to the models which we can construct geometrically. This situation is analogous to that which arises in Schild's work, wherein he shows that the velocities (other than zero) associated with integral time lines are very high, the smallest velocity being $\frac{1}{2}\sqrt{3} = 0.866$ time the velocity of light. The distances between individual lattice-points in our model (τ -time) correspond to the relative velocities in the uniformly expanding model (t -time); and the ratios of the former to the radius of hyperbolic space correspond to the ratios of the latter to the velocity of light and hence to the red-shifts. Our problem may therefore be expressed in the more picturesque form: what will be the order of magnitude of the smallest (non-zero) red-shifts in any hyperbolic model of discrete nebulae (or clusters) obeying the cosmological principle?

Although our primary concern is the construction of homogeneous honeycombs in hyperbolic space corresponding to Milne's problem as originally formulated, we shall also consider the same problem in spherical space, for this corresponds to the analogous problem in an Einstein-Eddington universe. There again we seek the ratio of the shortest distance between lattice-points and the radius of curvature; but as the total volume of space is now finite, we can express this problem in the more arresting form: how many discrete nebulae (or clusters), obeying the cosmological principle, can we pack into a spherical space? We shall see that the answer is exactly 14,400.

2. REGULAR AND QUASI-REGULAR TESSELLATIONS

It is an important property of Euclidean and non-Euclidean spaces that they are *homogeneous* and *isotropic*: there is a continuous group of congruent transformations that is transitive on the points, and there is a subgroup that is transitive on the

directions from any one point. It is interesting to see how nearly a discrete distribution of points can share these properties of the whole space.

Following Milne, we shall describe some distributions admitting a discrete group transitive on the points, with a finite (but 'not too small') subgroup transitive on the nearest neighbours of one point. For this purpose we consider the vertices of certain tessellations and honeycombs (Coxeter 1948, chap. IV). We begin in two dimensions, before working up to three.

In the Euclidean plane, the angle of a regular p -gon, say $\{p\}$, is $(1 - 2/p)\pi$; hence q equal $\{p\}$'s of any size will fit together round a common vertex if

$$1 - \frac{2}{p} = \frac{2}{q}, \quad \text{i.e.} \quad (p-2)(q-2) = 4.$$

We thus obtain the three regular tessellations $\{p, q\}$, namely,

$$\{4, 4\}, \quad \{3, 6\}, \quad \{6, 3\}. \quad (2.1)$$

The angle of a *spherical* $\{p\}$ is *greater* than $(1 - 2/p)\pi$, and gradually increases from this value to π when the circum-radius increases from 0 to $\frac{1}{2}\pi$. Hence, if

$$(p-2)(q-2) < 4,$$

we can adjust the size of the polygon so that the angle is exactly $2\pi/q$. Then q such $\{p\}$'s will fit together round a common vertex, and we have the spherical tessellations $\{p, q\}$, namely,

$$\{3, 3\}, \quad \{3, 4\}, \quad \{4, 3\}, \quad \{3, 5\}, \quad \{5, 3\},$$

corresponding to the five Platonic solids (for which we use the same symbols; see Coxeter 1948, pp. 5, 64).

In the *hyperbolic* plane, the angle of a $\{p\}$ is *less* than $(1 - 2/p)\pi$, and gradually decreases from this value to 0 when the circum-radius increases from 0 to ∞ . Hence, if

$$(p-2)(q-2) > 4,$$

we can adjust the size of the polygon so that the angle is exactly $2\pi/q$. Then q such $\{p\}$'s will fit together round a common vertex, and we can continue to add further $\{p\}$'s indefinitely. In this manner we construct a *hyperbolic tessellation* $\{p, q\}$, which is an infinite collection of regular p -gons filling the whole hyperbolic plane (Schlegel 1883, p. 360).

We can thus define a Euclidean or non-Euclidean tessellation $\{p, q\}$ for all integers p and q greater than 2. The centres of the $\{p\}$'s of $\{p, q\}$ are the vertices of the reciprocal tessellation (or 'dual map') $\{q, p\}$. Corresponding edges of $\{p, q\}$ and $\{q, p\}$ cross each other at their common mid-point, and such points are the vertices of another tessellation, say $\left\{ \begin{smallmatrix} p \\ q \end{smallmatrix} \right\}$, consisting of $\{p\}$'s and $\{q\}$'s, two of each at a vertex, arranged alternately. In particular, $\left\{ \begin{smallmatrix} p \\ p \end{smallmatrix} \right\}$ is the same as $\{p, 4\}$. (For drawings of portions of $\left\{ \begin{smallmatrix} 5 \\ 5 \end{smallmatrix} \right\}$, $\left\{ \begin{smallmatrix} 4 \\ 5 \end{smallmatrix} \right\}$, $\left\{ \begin{smallmatrix} 4 \\ 6 \end{smallmatrix} \right\}$, $\left\{ \begin{smallmatrix} 3 \\ 7 \end{smallmatrix} \right\}$, see Coxeter 1939, pp. 135, 136, 137, 139.)

The general *quasi-regular* tessellation consists of regular polygons of two kinds, each entirely surrounded by the other kind, say $\{p\}$'s and $\{q\}$'s, r of each at a vertex ($r \geq 2$). The centres of two adjacent polygons form, with either of their common vertices, a *characteristic triangle*, whose sides are lines of symmetry. Since the angles of this triangle are π/p , π/q , π/r , the tessellation is spherical or Euclidean or hyperbolic according as

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > \text{ or } = \text{ or } < 1.$$

Conversely, given such a triangle, the reflexions in its sides generate a group (e.g. Coxeter 1939, pp. 126, 127), and the transforms of one vertex of the triangle are the vertices of the tessellation (see Coxeter 1940, p. 391, or 1948, p. 66). The distance 2ϕ between neighbouring points, being twice an altitude of the characteristic triangle, is given by

$$\left. \begin{array}{l} \cos \phi \\ \text{or } \cosh \phi \end{array} \right\} = \sqrt{\left(\cos^2 \frac{\pi}{p} + \cos^2 \frac{\pi}{q} + 2 \cos \frac{\pi}{p} \cos \frac{\pi}{q} \cos \frac{\pi}{r} \right)} / \sin \frac{\pi}{r}$$

(except in the Euclidean case, when it is of course arbitrary). Such groups were first used by Klein (1879).

When $r = 2$ the tessellation is $\left\{ \begin{smallmatrix} p \\ q \end{smallmatrix} \right\}$, and when $p = 2$ it is $\{q, r\}$ (if we agree to disregard the digons, $\{2\}$, which collapse to form edges). In the latter case the three vertices of the characteristic triangle are a vertex of the tessellation, the mid-point of an edge, and the centre of a polygon. The quasi-regular tessellation $\left\{ \begin{smallmatrix} 3 \\ 7 \end{smallmatrix} \right\}$ has $2\phi = 0.492 \dots$ (if we take the curvature of the hyperbolic plane to be -1).

In the spherical case the area of the characteristic triangle is $\left(\frac{1}{p} + \frac{1}{q} + \frac{1}{r} - 1 \right) \pi$; therefore the order of the group is

$$4 / \left(\frac{1}{p} + \frac{1}{q} + \frac{1}{r} - 1 \right),$$

and the number of vertices of the tessellation is this order divided by $2r$. Setting r or p equal to 2, we infer that the numbers of vertices of $\left\{ \begin{smallmatrix} p \\ q \end{smallmatrix} \right\}$ and $\{q, r\}$ are respectively

$$\frac{2pq}{4 - (p-2)(q-2)} \quad \text{and} \quad \frac{4q}{4 - (q-2)(r-2)}. \quad (2.2)$$

In the hyperbolic case such distributions of points remain interesting when $q = \infty$. In fact, we shall find that the hyperbolic tessellations

$$\left\{ \begin{smallmatrix} 3 \\ \infty \end{smallmatrix} \right\}, \quad \left\{ \begin{smallmatrix} 4 \\ \infty \end{smallmatrix} \right\}, \quad \{\infty, 3\}, \quad \{\infty, 4\}$$

are especially amenable to analytic treatment. The apeirogon $\{\infty\}$, being the limiting form approached by $\{q\}$ when q increases, is inscriptible in a *horocycle* (i.e. a circle of infinite radius).

We use projective co-ordinates, with the absolute conic represented by an equation $\Omega = 0$ of the second degree. Since a horocycle has four coincident intersections with this conic (Coxeter 1942, p. 218), its equation takes the form

$$t^2 = c\Omega,$$

where $t = 0$ is the common tangent. If the co-ordinates are normalized so that $\Omega = 1$ and $t > 0$, the equation for the horocycle becomes simply

$$t = \sqrt{c}.$$

To obtain co-ordinates for the vertices of one of the regular tessellations, we take $\Omega = x_0^2 - x_1^2 - x_2^2$ and consider the points whose co-ordinates are the integral solutions of the Diophantine equation

$$x_0^2 - x_1^2 - x_2^2 = 1. \quad (2.3)$$

The points nearest to $(1, 0, 0)$ are $(3, \pm 2, \pm 2)$, forming a square. Two adjacent vertices $(3, 2, \pm 2)$ of this square belong to an $\{\infty\}$ inscribed in the horocycle

$$x_0 - x_1 = 1$$

(Coxeter 1942, p. 250), namely, $(2n^2 + 1, 2n^2, 2n)$ for all integers n . The four sides of the square provide four such horocyclic apeirogons,

$$(2n^2 + 1, 2n^2, 2n), \quad (2n^2 + 1, -2n^2, 2n),$$

$$(2n^2 + 1, 2n, 2n^2), \quad (2n^2 + 1, 2n, -2n^2),$$

surrounding the common vertex $(1, 0, 0)$. This is the beginning of a tessellation of $\{\infty\}$'s, 4 at a vertex, namely,

$$\{\infty, 4\}.$$

The characteristic triangle is formed by the vertex $(1, 0, 0)$, the mid-edge point $(2, 1, 1)$, and the face-centre $(1, 1, 0)$ (which is the point at infinity on the horocycle). Thus its sides are

$$x_0 = x_1 + x_2, \quad x_2 = 0, \quad x_1 = x_2,$$

and the generating reflexions are

$$(A) \quad x'_i = x_i + 2(x_0 - x_1 - x_2),$$

$$(B) \quad \text{the reversal of sign of } x_2,$$

$$(C) \quad \text{the transposition } (x_1 x_2).$$

Since these transformations are automorphs of the quadratic form Ω , every vertex of the $\{\infty, 4\}$ provides a solution of the equation (2.3).

Conversely, every solution represents a vertex. To prove this, let (x_0, x_1, x_2) be *any* solution. It represents some point, and since the co-ordinates are homogeneous there will be no loss of generality in assuming that $x_0 > 0$.

If $x_0 > 1$, a suitable combination of the above three transformations will yield another solution with a smaller (positive) x_0 . In fact, the condition $x_0 > 1$ implies $x_1 \neq 0$ and $x_2 \neq 0$, and the trivial transformations B and C enable us to assume $x_1 \geq 1, x_2 \geq 1$, so that

$$x_0 = \sqrt{(x_1^2 + x_2^2 + 1)} < x_1 + x_2 \leq \sqrt{2(x_1^2 + x_2^2)} < (\sqrt{2})x_0 < \frac{3}{2}x_0,$$

whence

$$x_0 - x_1 - x_2 < 0 \quad \text{and} \quad x_0 + 2(x_0 - x_1 - x_2) > 0.$$

Thus the transformation A decreases each co-ordinate while leaving x_0 positive. By repeated application we must eventually reach a solution with $x_0 = 1$, namely, $(1, 0, 0)$.

Reversing the procedure, we have a sequence of symmetry operations leading from $(1, 0, 0)$ to the given solution (x_0, x_1, x_2) . Thus we have proved that *the vertices of $\{\infty, 4\}$ are represented by the integral solutions of*

$$x_0^2 - x_1^2 - x_2^2 = 1.$$

Applying the same transformations to the point $(2, 1, 1)$, which is midway between $(1, 0, 0)$ and $(3, 2, 2)$, we see that *the vertices of $\left\{ \begin{smallmatrix} 4 \\ \infty \end{smallmatrix} \right\}$ are represented by the integral solutions of*

$$x_0^2 - x_1^2 - x_2^2 = 2.$$

The expression $x_0^2 - x_1^2 - x_2^2$ for Ω implies that we are using a self-polar triangle of reference. We might use instead a triangle of reference *inscribed* in the absolute conic, so that $\Omega = yz + zx + xy$. Reasoning similar to the above enables us to see that the integral solutions of the equation

$$yz + zx + xy = 1$$

represent the vertices of $\left\{ \begin{smallmatrix} 3 \\ \infty \end{smallmatrix} \right\}$. Typical faces are the triangle

$$(0, 1, 1), \quad (1, 0, 1), \quad (1, 1, 0)$$

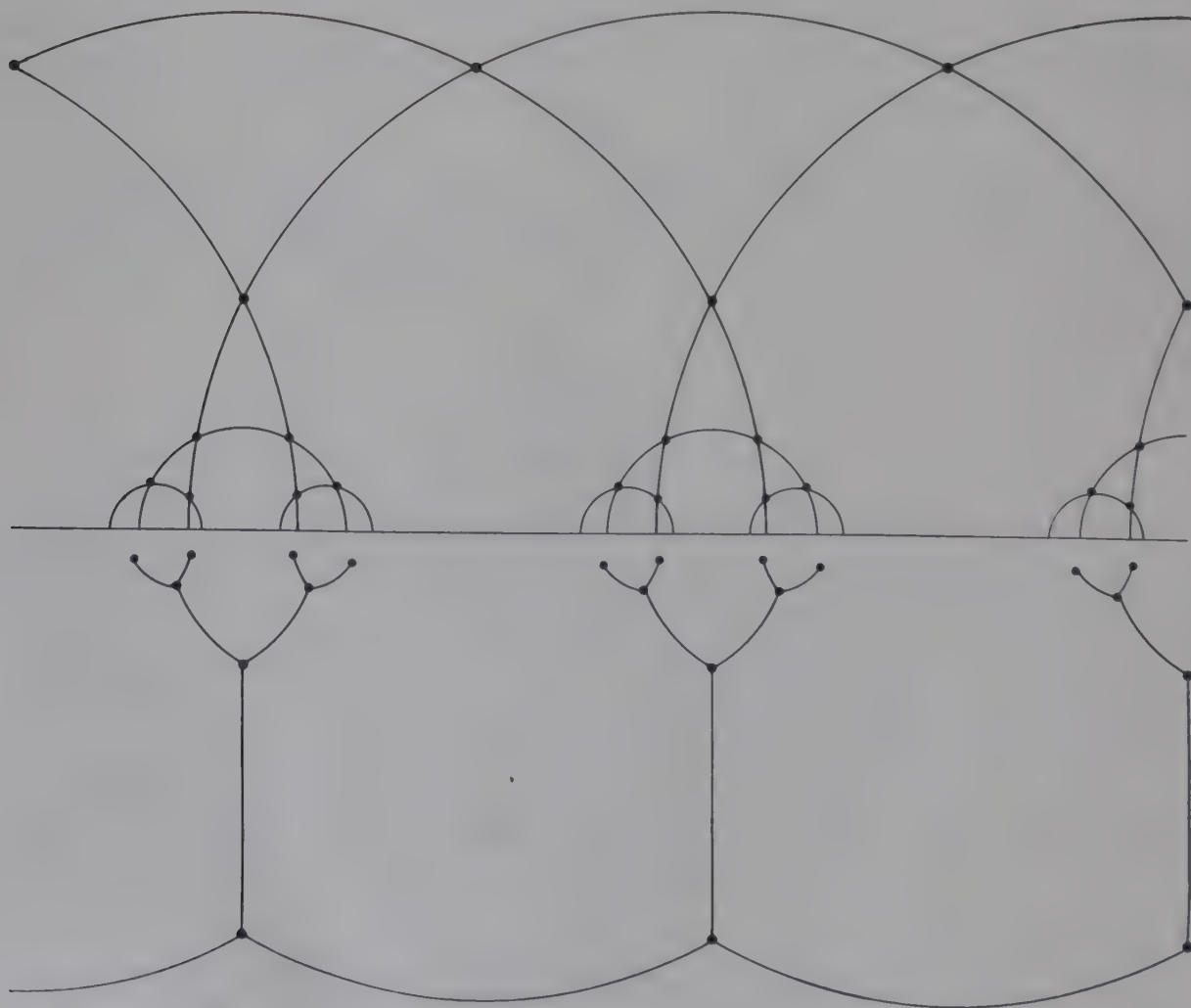


FIGURE 1

and an apeirogon $(x, 1-x, 1-x+x^2)$ inscribed in the horocycle $x+y=1$. The characteristic triangle has sides $x=0, \quad x=y, \quad y=z$.

The reflexion in the first side is

$$x' = -x, \quad y' = y + 2x, \quad z' = z + 2x,$$

while the other two reflexions are mere transpositions; thus the number of odd co-ordinates is invariant. It follows that the *odd* solutions of

$$yz + zx + xy = 3$$

represent the vertices of $\{\infty, 3\}$, with a typical face

$$(1-2n, 1+2n, 1+2n^2)$$

inscribed in the horocycle $x+y=2$.

Fragments of $\left\{\frac{3}{\infty}\right\}$ and $\{\infty, 3\}$ are represented conformally in figure 1 (adapted from figure 151 of Fricke & Klein 1897). In this 'Poincaré map', lines are represented by semicircles (and lines) orthogonal to the thin central line, and horocycles by circles touching it (and lines parallel to it).

3. REGULAR AND QUASI-REGULAR HONEYCOMBS

In Euclidean 3-space, equal cubes $\{4, 3\}$ can be fitted together, 4 round each edge, to form the regular honeycomb $\{4, 3, 4\}$. The arrangement of cells round a common vertex corresponds to the arrangement of faces of a solid called the *vertex figure*, in the present case an octahedron $\{3, 4\}$. The *Schläfli symbol* $\{4, 3, 4\}$ is obtained by telescoping the symbols for the cell and vertex figure.

In spherical 3-space (i.e. on the hypersurface of a sphere in Euclidean 4-space) the dihedral angle of a regular polyhedron $\{p, q\}$ is greater than that of the corresponding Euclidean polyhedron, and increases to π when the circum-radius increases to $\frac{1}{2}\pi$. It is thus possible to fit, round a common edge, 3 or 4 or 5 tetrahedra, or 3 hexahedra, or 3 octahedra, or 3 dodecahedra, so as to construct the six spherical honeycombs

$$\{3, 3, 3\}, \quad \{3, 3, 4\}, \quad \{3, 3, 5\}, \quad \{4, 3, 3\}, \quad \{3, 4, 3\}, \quad \{5, 3, 3\}$$

(Coxeter 1948, pp. 136–137).

In hyperbolic 3-space, the dihedral angle of $\{p, q\}$ is less than that of the corresponding Euclidean polyhedron, and decreases to $(1-2/q)\pi$ when the circum-radius increases to ∞ . Thus it is possible to fit, round a common edge, 4 or 5 dodecahedra, or 5 hexahedra or 3 icosahedra, forming the hyperbolic honeycombs

$$\{5, 3, 4\}, \quad \{5, 3, 5\}, \quad \{4, 3, 5\}, \quad \{3, 5, 3\}$$

(Schlegel 1883, p. 444; Stringham 1880, pp. 7, 12 and errata; Sommerville 1924).

We can thus define a honeycomb $\{p, q, r\}$ whenever the cell $\{p, q\}$ and vertex figure $\{q, r\}$ are Platonic solids; and this honeycomb is spherical or Euclidean or hyperbolic according as

$$\sin \frac{\pi}{p} \sin \frac{\pi}{r} > \text{ or } = \text{ or } < \cos \frac{\pi}{q}$$

(Coxeter 1948, p. 135; Sommerville 1929, pp. 168–191).

Since we are primarily concerned with arrangements of *points*, we shall extend these ideas so as to allow the cell $\{p, q\}$ of a hyperbolic honeycomb to be of one of the Euclidean types (2·1), with its vertices lying on a *horosphere* (i.e. a sphere of infinite radius; see Coxeter 1942, pp. 220, 251). The extra honeycombs thus admitted are

$$\{4, 4, 3\}, \quad \{6, 3, 3\}, \quad \{6, 3, 4\}, \quad \{6, 3, 5\}.$$

In other words, we consider honeycombs $\{p, q, r\}$ satisfying

$$(p-2)(q-2) \leq 4 \quad \text{and} \quad (q-2)(r-2) < 4.$$

This does not look symmetrical; but if we allowed $(q-2)(r-2) = 4$, the *vertices* would be at infinity, instead of merely the centres of the cells.

Besides these regular honeycombs, there are some 'quasi-regular' ones that can be constructed as follows. We recall that four of the eight vertices of the cube $\{4, 3\}$ belong to a tetrahedron $\{3, 3\}$; more generally, alternate vertices of $\{4, q\}$ belong to $\{q, q\}$. In a convenient notation (with *h* standing for 'half' or 'hemi') we express this relation as

$$h\{4, q\} = \{q, q\}.$$

It follows that alternate vertices of $\{4, q, r\}$ belong to a honeycomb whose cells are $\{q, q\}$ and $\{q, r\}$, say

$$h\{4, q, r\} = \left\{ q, \begin{smallmatrix} q \\ r \end{smallmatrix} \right\}$$

(Coxeter 1948, pp. 70, 155). We thus obtain the quasi-regular honeycombs

$$h\{4, 3, 4\}, \quad h\{4, 3, 5\}, \quad h\{4, 4, 3\},$$

of which the first is Euclidean while the other two are hyperbolic. The Euclidean honeycomb $h\{4, 3, 4\}$ is particularly interesting as its vertices form the 'face-centred cubic lattice', i.e. they are the centres of the spheres in 'cubic close-packing'.

Another way to obtain the same figures is to consider a honeycomb of $\{p, q\}$'s and $\{p, r\}$'s, each cell of either kind being surrounded entirely by the other kind, so that each edge belongs to (say), $s\{p, q\}$'s and $s\{p, r\}$'s, arranged alternately. Then the vertex figure is a finite polyhedron whose faces are $\{q\}$'s and $\{r\}$'s, s of each at a vertex; and we have seen that this is either $\{q, 2s\}$ (with $q = r$) or $\{q, s\}$ (with $r = 2$) or $\left\{ \begin{smallmatrix} q \\ r \end{smallmatrix} \right\}$ (with $s = 2$). Hence the honeycomb is either

$$\{p, q, 2s\} \quad \text{or} \quad \{p, q, s\} \quad \text{or} \quad \left\{ p, \begin{smallmatrix} q \\ r \end{smallmatrix} \right\}.$$

In the last case, where the symbol means that the two kinds of cell are $\{p, q\}$ and $\{p, r\}$ while the vertex figure is $\left\{ \begin{smallmatrix} q \\ r \end{smallmatrix} \right\}$, we have the restrictions

$$(p-2)(q-2) \leq 4, \quad (p-2)(r-2) \leq 4, \quad (q-2)(r-2) < 4,$$

which imply that p, q, r are not all different; so the only possibilities are

$$\left\{ p, \begin{smallmatrix} q \\ r \end{smallmatrix} \right\} = \{p, q, 4\} \quad \text{and} \quad \left\{ \begin{smallmatrix} q \\ r \end{smallmatrix} \right\} = h\{4, q, r\}.$$

In table 1, N is the number of vertices nearest to any given vertex, or the number of edges at a vertex, or the number of vertices of the vertex figure; and 2ϕ is the edge-length, i.e. the distance between neighbouring vertices. Since the vertex figure is $\{q, r\}$ or $\begin{Bmatrix} q \\ r \end{Bmatrix}$, (2.2) shows that the values of N for $\{p, q, r\}$ and $h\{4, q, r\}$ are respectively

$$\frac{4q}{4-(q-2)(r-2)} \quad \text{and} \quad \frac{2qr}{4-(q-2)(r-2)}.$$

The edge-length of the regular honeycomb $\{p, q, r\}$ is given in the spherical case by the formula

$$\cos \phi = \cos \frac{\pi}{p} \sin \frac{\pi}{r} / \sin \frac{\pi}{h}, \quad \text{where} \quad h = \sqrt{\left\{ \frac{2(q+r)+7qr}{2(q+r)-qr} \right\} - 1}$$

(Coxeter 1948, pp. 139, 15). In the hyperbolic case we have to solve a hyperbolic triangle instead of a spherical triangle, and the result is similarly

$$\cosh \phi = \cos \frac{\pi}{p} \sin \frac{\pi}{r} / \sin \frac{\pi}{h}.$$

The edge-length 2ϕ of $h\{4, q, r\}$ is expressible in terms of the edge-length $2\phi'$ of $\{4, q, r\}$ itself by the relation

$$\cos^2 \phi = \cos 2\phi' = 2 \cos^2 \phi' - 1 \quad (\text{spherical})$$

$$\text{or} \quad \cosh^2 \phi = \cosh 2\phi' = 2 \cosh^2 \phi' - 1 \quad (\text{hyperbolic}).$$

For, 2ϕ is the diagonal of a $\{4\}$ of side $2\phi'$, which is a face of $\{4, q, r\}$; and one-quarter of this face is an isosceles right-angled triangle of sides $\phi, \phi, 2\phi'$, so that

$$\cos 2\phi' = \cos^2 \phi \quad \text{or} \quad \cosh 2\phi' = \cosh^2 \phi$$

(Coxeter 1942, p. 242).

TABLE 1. TABLE OF REGULAR AND QUASI-REGULAR HONEYCOMBS*

honeycomb	N	h	$\cos^2 \pi/p$	$\sin^2 \pi/r$	$\sin^2 \pi/h$	$\cos^2 \phi$	$\cosh^2 \phi$	2ϕ
$\{3, 3, 3\}$	4	4	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{8}$	—	1.8235
$\{3, 3, 4\}$	6	6	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	—	1.5708
$\{4, 3, 3\}$	4	4	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	—	1.0472
$\{3, 4, 3\}$	8	6	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	—	1.0472
$\{3, 3, 5\}$	12	10	$\frac{1}{4}$	$\frac{1}{4}\sqrt{5}\tau^{-1}$	$\frac{1}{4}\tau^{-2}$	$\frac{1}{4}\sqrt{5}\tau$	—	0.6283
$\{5, 3, 3\}$	4	4	$\frac{1}{4}\tau^2$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{8}\tau^2$	—	0.2709
$\{4, 3, 4\}$	6	6	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	1	1	0
$h\{4, 3, 4\}$	12	—	—	—	—	1	1	0
$\{6, 3, 3\}$	4	4	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	—	$\frac{9}{8}$	0.6931†
$\{5, 3, 4\}$	6	6	$\frac{1}{4}\tau^2$	$\frac{1}{2}$	$\frac{1}{4}$	—	$\frac{1}{2}\tau^2$	1.0613
$\{6, 3, 4\}$	6	6	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	—	$\frac{3}{2}$	1.3170
$\{4, 4, 3\}$	8	6	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$	—	$\frac{3}{2}$	1.3170
$h\{4, 4, 3\}$	12	—	—	—	—	—	2	1.7627
$\{4, 3, 5\}$	12	10	$\frac{1}{2}$	$\frac{1}{4}\sqrt{5}\tau^{-1}$	$\frac{1}{4}\tau^{-2}$	—	$\frac{1}{2}\sqrt{5}\tau$	1.6169
$h\{4, 3, 5\}$	30	—	—	—	—	—	τ^2	2.1226
$\{3, 5, 3\}$	20	10	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{4}\tau^{-2}$	—	$\frac{3}{4}\tau^2$	1.7366
$\{5, 3, 5\}$	12	10	$\frac{1}{4}\tau^2$	$\frac{1}{4}\sqrt{5}\tau^{-1}$	$\frac{1}{4}\tau^{-2}$	—	$\frac{1}{4}\sqrt{5}\tau^3$	1.9928
$\{6, 3, 5\}$	12	10	$\frac{3}{4}$	$\frac{1}{4}\sqrt{5}\tau^{-1}$	$\frac{1}{4}\tau^{-2}$	—	$\frac{3}{4}\sqrt{5}\tau$	2.1678

* We use the abbreviation $\tau = 2 \cos \frac{1}{5}\pi = \frac{1}{2}(\sqrt{5} + 1)$.

† $\log 2$.

We observe that the regular spherical honeycomb with smallest ϕ is $\{5, 3, 3\}$, which has 600 vertices and 120 dodecahedral cells (Coxeter 1948, p. 153), and that the regular hyperbolic honeycomb with smallest ϕ is $\{6, 3, 3\}$, whose cells are horospherical. No convenient co-ordinates for this hyperbolic honeycomb have been found; but $\{4, 4, 3\}$ has likewise horospherical cells, and can be constructed as follows.

Let the integral solutions of the equation

$$x_0^2 - x_1^2 - x_2^2 - x_3^2 = 1 \quad (3.1)$$

be regarded as homogeneous co-ordinates for points (x_0, x_1, x_2, x_3) in hyperbolic space with absolute quadric

$$x_0^2 - x_1^2 - x_2^2 - x_3^2 = 0.$$

The points nearest to $(1, 0, 0, 0)$ are $(2, \pm 1, \pm 1, \pm 1)$, which we recognize as the vertices of a cube (or perhaps we should rather say, the space being hyperbolic, a 'regular hexahedron'). The four vertices $(2, 1, \pm 1, \pm 1)$ of a square face of this cube belong to a tessellation $\{4, 4\}$ on the horosphere $x_0 - x_1 = 1$, namely,

$$(m^2 + n^2 + 1, m^2 + n^2, m + n, m - n)$$

for all integers m and n . Thus we have the beginning of a honeycomb of $\{4, 4\}$'s with vertex figure $\{4, 3\}$, namely,

$$\{4, 4, 3\}.$$

The characteristic tetrahedron is formed by the vertex $(1, 0, 0, 0)$, the mid-edge point $(3, 1, 1, 1)$, the face-centre $(2, 1, 1, 0)$, and the cell-centre $(1, 1, 0, 0)$ (at infinity). Thus its face-planes (omitting these points one by one) are

$$x_0 = x_1 + x_2 + x_3, \quad x_3 = 0, \quad x_2 = x_3, \quad x_1 = x_2,$$

and the four generating reflexions are:

- (A) $x'_i = x_i + (x_0 - x_1 - x_2 - x_3)$,
- (B) the reversal of sign of x_3 ,
- (C) the transposition $(x_2 x_3)$,
- (D) the transposition $(x_1 x_2)$.

We verify that A is a Lorentz transformation by observing that

$$(2x_0 - x_1 - x_2 - x_3)^2 - (x_0 - x_2 - x_3)^2 - (x_0 - x_1 - x_3)^2 - (x_0 - x_1 - x_2)^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2.$$

To make sure that *every* solution of (3.1) represents a vertex of $\{4, 4, 3\}$, let (x_0, x_1, x_2, x_3) be any solution. Since the co-ordinates are homogeneous there will be no loss of generality in assuming that $x_0 > 0$.

If $x_0 > 1$, a suitable combination of the above four transformations will yield another solution with a smaller x_0 . In fact, after using B, C and D (if necessary) to make every $x_i \geq 0$, we have

$$x_0 = \sqrt{(x_1^2 + x_2^2 + x_3^2 + 1)} < x_1 + x_2 + x_3 + 1 \leq \sqrt{4(x_1^2 + x_2^2 + x_3^2 + 1)} = 2x_0,$$

whence

$$x_0 - x_1 - x_2 - x_3 < 1 \quad \text{and} \quad 2x_0 - x_1 - x_2 - x_3 \geq 1.$$

But $x_0 - x_1 - x_2 - x_3$, being odd, cannot vanish; therefore it must be negative. Thus A decreases each co-ordinate while leaving x_0 positive. By repeated application we eventually reach a solution with $x_0 = 1$, namely, $(1, 0, 0, 0)$.

Reversing the procedure, we have a sequence of symmetry operations leading from $(1, 0, 0, 0)$ to the given solution (x_0, x_1, x_2, x_3) . Thus we have proved that *the vertices of $\{4, 4, 3\}$ are represented by the integral solutions of*

$$x_0^2 - x_1^2 - x_2^2 - x_3^2 = 1.$$

Incidentally, it follows that the operations A, B, C, D generate the group of all Lorentz transformations with integral coefficients (Schild 1949, p. 39), and that this infinite discrete group can be abstractly defined by the generating relations

$$A^2 = B^2 = C^2 = D^2 = (AC)^2 = (AD)^2 = (BD)^2 = (AB)^4 = (BC)^4 = (CD)^3 = 1$$

(Coxeter 1948, p. 80).

Since A increases or decreases each co-ordinate by the same amount, every solution has either x_0 even and the other three odd, or vice versa. This distinction enables us to regard the vertices of $\{4, 4, 3\}$ as belonging to two types. For any edge

$$(x_0, x_1, x_2, x_3)(y_0, y_1, y_2, y_3)$$

we have $x_0y_0 - x_1y_1 - x_2y_2 - x_3y_3 = \cosh 2\phi = 2$; therefore every edge joins vertices of opposite types, and the two types belong to the two $h\{4, 4, 3\}$'s that can be inscribed in $\{4, 4, 3\}$.

In the $h\{4, 4, 3\}$ formed by the solutions of (3.1) with x_0 odd and the other x 's even, the 12 vertices nearest to $(1, 0, 0, 0)$ are

$$(3, 0, \pm 2, \pm 2), \quad (3, \pm 2, 0, \pm 2), \quad (3, \pm 2, \pm 2, 0),$$

which we recognize as the vertices of a cuboctahedron (Coxeter 1948, p. 52). Comparing (3.1) with (2.3), we see that the section of $h\{4, 4, 3\}$ by the plane $x_3 = 0$ is $\{\infty, 4\}$.

4. LOCALLY ANISOTROPIC HONEYCOMBS

The honeycombs considered above are not merely homogeneous but 'almost isotropic'; the group that is transitive on the vertices has a polyhedral subgroup transitive on the edges at one vertex. We can reasonably expect to find homogeneous honeycombs of smaller edge-length (or 'finer mesh') by relaxing the latter condition, asking only for a kind of 'isotropy in the large' to exclude certain distributions of points that are less than three-dimensional, like an unbounded chess board drawn on a Clifford surface (in spherical space) or on a horosphere (in hyperbolic space).

To obtain a homogeneous distribution of points we merely have to take the transforms of an arbitrary point under a discrete group of congruent transformations. The enumeration of such groups in Euclidean space was one of the most spectacular cases of independent discovery by several investigators; for the number 230 was obtained by Federov in Russia, Schoenflies in Germany, and Barlow in England, all between 1885 and 1895. It is interesting to record that the corresponding enumeration for elliptic space was carried out during the same period by Goursat (1889). This was extended to spherical space by Threlfall & Seifert (1931). But the *hyperbolic* space groups have never been enumerated.

Being interested in 'close-grained' distributions, we seek groups whose fundamental regions are as small as possible. Analogy with the spherical case suggests that such a group will be closely related to one generated by reflexions in the

four faces of a tetrahedron. It may be just a group generated by reflexions, in which case the tetrahedron is its fundamental region; but, if the tetrahedron is symmetrical, we can obtain a larger group by adjoining the symmetry operations of the tetrahedron itself. If such a symmetry operation is again a reflexion, we merely have another reflexion group with a smaller fundamental region; so the significant cases arise when the tetrahedron is symmetrical by rotation, namely, either by one half-turn or by two commutative half-turns generating a 4-group.

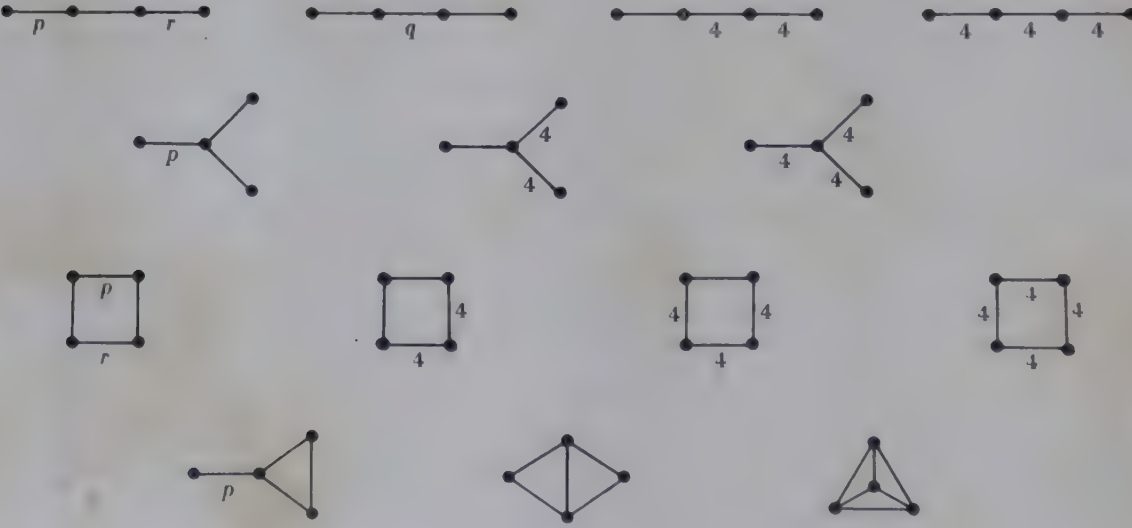
The dihedral angles of the fundamental region of a group generated by reflexions are submultiples of π (Coxeter 1948, pp. 76–80). When the fundamental region is a tetrahedron, it is conveniently represented by a graph (Coxeter 1948, pp. 84, 191). There are four nodes, representing the bounding planes, and these are linked by branches indicating which pairs of planes are not perpendicular. A branch marked p indicates a dihedral angle π/p ; but the mark 3, which otherwise would occur tiresomely often, is omitted. By removing from the graph each node in turn (along with any branches belonging to that node) we derive symbols for the four trihedral angles. On a sphere drawn round a proper vertex of the tetrahedron, the trihedral angle determines a spherical triangle whose angles are submultiples of π , namely, one of the following:



(Coxeter 1948, pp. 82, 85). In the hyperbolic case we allow one or more vertices to be at infinity, and we take such a point to be the centre of a horosphere on which the three planes cut out a horospherical triangle. The corresponding part of the graph is then one of



Accordingly, we seek graphs of four nodes, so arranged as to leave one of these subgraphs when any one of the nodes is removed. Setting aside disconnected graphs (representing reducible groups which would be ruled out by the condition of ‘isotropy in the large’), we soon find the following list of possibilities:



Here p and r take the values 3, 4, 5, 6 independently ($p \leq r$), and q takes the values 4, 5, 6.

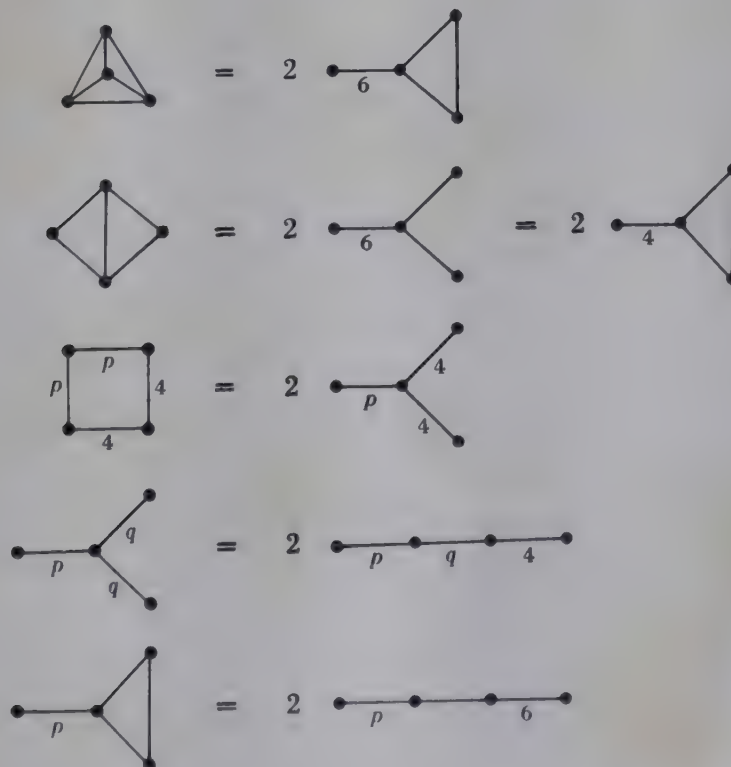
Five of these tetrahedra are known to be spherical, and three Euclidean (Coxeter 1948, p. 297). The remaining 32, having smaller angles, must be hyperbolic.

Each tetrahedron will serve as fundamental region for a group generated by reflexions, and the transforms of any point will provide a homogeneous distribution. If the chosen point is at a vertex of the tetrahedron, the resulting honeycomb is symbolized by drawing a ring round the corresponding node of the graph (Coxeter 1948, p. 87); e.g. the honeycombs $\{p, q, r\}$ and $\left\{p, \frac{q}{r}\right\}$ are

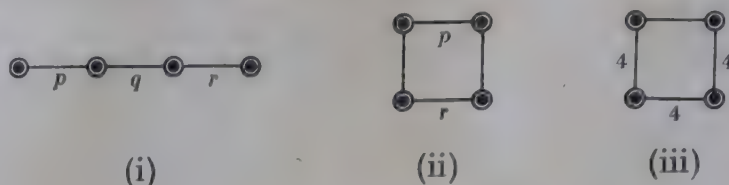


Similarly, by drawing rings round two, three, or all four nodes, we symbolize the honeycombs obtained by choosing a point on an edge of the tetrahedron, or on a face, or quite inside (Coxeter 1940, pp. 402–404). In the last case, where every node is ringed, each vertex of the honeycomb has four neighbours, derived by the four generating reflexions. In order to make these four neighbours equidistant from the first vertex, we take the latter at the *in-centre* of the tetrahedron. (We could make *some* edges shorter by selecting a different position, but then the shortest edges would not form a connected graph; in other words, the vertices would occur in widely separated clusters.)

This procedure is known as ‘Wythoff’s construction’ (Coxeter 1935). Since one fundamental region may yield as many as $4 + 6 + 4 + 1 = 15$ different kinds of honeycomb, the total number is quite large. However, a certain amount of duplication arises when a tetrahedron has a plane of symmetry by which it can be cut in half, the half being another tetrahedron in the list. These cases are as follows (cf. Coxeter 1948, pp. 200, 281):



Thus the honeycomb of shortest edge derivable by Wythoff's construction must occur among



Since the vertices of the honeycomb are the centres of spheres inscribed in the tetrahedra, we seek the *in-radius* ϕ of such a tetrahedron. This is most readily computed by regarding the tetrahedron as tetrahedron of reference for homogeneous point-co-ordinates (x_0, x_1, x_2, x_3) and plane-co-ordinates $[X_0, X_1, X_2, X_3]$, with the absolute polarity in its general form

$$X_\mu = \sum c_{\mu\nu} x_\nu, \quad x_\mu = \sum C_{\mu\nu} X_\nu \quad (\mu = 0, 1, 2, 3),$$

where $c_{\mu\nu} = c_{\nu\mu}$, $\sum c_{\lambda\nu} C_{\mu\nu} = \delta_{\lambda\mu}$, and the Σ implies summation over the repeated suffix (cf. Coxeter 1942, chap. XII). Here $\|c_{\mu\nu}\|$ and $\|C_{\mu\nu}\|$ are inverse matrices, so that the determinant

$$\Gamma = \det(C_{\mu\nu})$$

has $\Gamma c_{\mu\nu}$ as cofactor of its element $C_{\mu\nu}$.

In elliptic geometry, the dihedral angles $\Lambda_{\mu\nu}$ of the tetrahedron are given by

$$\cos \Lambda_{\mu\nu} = -C_{\mu\nu}/C_\mu C_\nu, \quad \text{where} \quad C_\mu = \sqrt{C_{\mu\mu}},$$

and the in-radius ϕ is given by

$$\sin \phi = (\Sigma \Sigma c_{\mu\nu} C_\mu C_\nu)^{-\frac{1}{2}}.$$

In hyperbolic geometry, where $C_{\mu\mu}$ is negative, we write $C_\mu = \sqrt{-C_{\mu\mu}}$ and find

$$\cos \Lambda_{\mu\nu} = C_{\mu\nu}/C_\mu C_\nu, \quad \sinh \phi = (\Sigma \Sigma c_{\mu\nu} C_\mu C_\nu)^{-\frac{1}{2}}.$$

Since a tetrahedron is determined by its six dihedral angles, we lose no generality by assuming $C_0 = C_1 = C_2 = C_3 = 1$, so that

$$C_{\mu\mu} = \pm 1, \quad C_{\mu\nu} = \mp \cos \Lambda_{\mu\nu} \quad (\mu \neq \nu).$$

Then we have simply $\sin \phi$ or $\sinh \phi = (\Sigma \Sigma c_{\mu\nu})^{-\frac{1}{2}}$.

Applying these formulae to the tetrahedron $q \begin{array}{c} \text{---} p \text{---} \\ | \quad | \\ \text{---} r \text{---} \end{array} s$, for which

$$\Lambda_{01} = \pi/p, \quad \Lambda_{12} = \pi/q, \quad \Lambda_{23} = \pi/r, \quad \Lambda_{30} = \pi/s, \quad \Lambda_{02} = \Lambda_{13} = \pi/2,$$

we have
$$\Gamma = \sin^2 \frac{\pi}{p} \sin^2 \frac{\pi}{r} + \sin^2 \frac{\pi}{q} \sin^2 \frac{\pi}{s} - 2 \cos \frac{\pi}{p} \cos \frac{\pi}{q} \cos \frac{\pi}{r} \cos \frac{\pi}{s} - 1,$$

$$\pm \Gamma c_{00} = \sin^2 \frac{\pi}{q} + \sin^2 \frac{\pi}{r} - 1,$$

$$\pm \Gamma c_{01} = \cos \frac{\pi}{p} \sin^2 \frac{\pi}{r} + \cos \frac{\pi}{q} \cos \frac{\pi}{r} \cos \frac{\pi}{s},$$

$$\pm \Gamma c_{02} = \cos \frac{\pi}{p} \cos \frac{\pi}{q} + \cos \frac{\pi}{r} \cos \frac{\pi}{s},$$

with the remaining cofactors given by simultaneous cyclic permutation of 0123 and $pqrs$. Hence

$$\begin{aligned} \Sigma \Sigma c_{\mu\nu} = \frac{2}{|\Gamma|} & \left\{ \sin^2 \frac{\pi}{p} \left(1 + \cos \frac{\pi}{r} \right) + \sin^2 \frac{\pi}{q} \left(1 + \cos \frac{\pi}{s} \right) + \sin^2 \frac{\pi}{r} \left(1 + \cos \frac{\pi}{p} \right) \right. \\ & + \sin^2 \frac{\pi}{s} \left(1 + \cos \frac{\pi}{q} \right) + \cos \frac{\pi}{q} \cos \frac{\pi}{r} \cos \frac{\pi}{s} \\ & + \cos \frac{\pi}{r} \cos \frac{\pi}{s} \cos \frac{\pi}{p} + \cos \frac{\pi}{s} \cos \frac{\pi}{p} \cos \frac{\pi}{q} + \cos \frac{\pi}{p} \cos \frac{\pi}{q} \cos \frac{\pi}{r} \\ & \left. + \cos \frac{\pi}{p} \cos \frac{\pi}{q} + \cos \frac{\pi}{q} \cos \frac{\pi}{r} + \cos \frac{\pi}{r} \cos \frac{\pi}{s} + \cos \frac{\pi}{s} \cos \frac{\pi}{p} - 2 \right\}. \end{aligned}$$

We now consider separately the three cases

$$(i) \ s = 2; \quad (ii) \ q = s = 3; \quad (iii) \ r = s = 4.$$

$$(i) \ \Gamma = \sin^2 \frac{\pi}{p} \sin^2 \frac{\pi}{r} - \cos^2 \frac{\pi}{q},$$

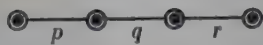
$$\Sigma \Sigma c_{\mu\nu} = \frac{2}{|\Gamma|} \left\{ \left(1 + \cos \frac{\pi}{p} \right) \left(1 + \cos \frac{\pi}{r} \right) \left(2 - \cos \frac{\pi}{p} + \cos \frac{\pi}{q} - \cos \frac{\pi}{r} \right) - \cos^2 \frac{\pi}{q} \right\}.$$

$$(ii) \ \Gamma = \sin^2 \frac{\pi}{p} \sin^2 \frac{\pi}{r} - \frac{1}{2} \cos \frac{\pi}{p} \cos \frac{\pi}{r} - \frac{7}{16},$$

$$\Sigma \Sigma c_{\mu\nu} = \frac{1}{2|\Gamma|} \left(3 - \cos \frac{\pi}{p} - \cos \frac{\pi}{r} \right) \left\{ 4 \left(1 + \cos \frac{\pi}{p} \right) \left(1 + \cos \frac{\pi}{r} \right) - 1 \right\}.$$

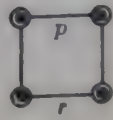
$$(iii) \ \Gamma = -\frac{1}{2} \left(\cos \frac{\pi}{p} + \cos \frac{\pi}{q} \right)^2,$$

$$\Sigma \Sigma c_{\mu\nu} = \frac{\sqrt{2}+1}{|\Gamma|} \left\{ \left(1 + \sqrt{2} \cos \frac{\pi}{p} \right) \left(1 + \sqrt{2} \cos \frac{\pi}{q} \right) + \sqrt{2} \left(1 - \cos^2 \frac{\pi}{p} - \cos^2 \frac{\pi}{q} \right) \right\}.$$

TABLE 2. TABLE OF HONEYCOMBS 

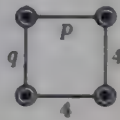
p	q	r	Γ	$\Sigma \Sigma c_{\mu\nu}$	2ϕ
3	3	3	$\frac{5}{16}$	20	0.4436
3	3	4	$\frac{1}{8}$	$32 + 12\sqrt{2}$	0.2848
3	4	3	$\frac{1}{16}$	$56 + 36\sqrt{2}$	0.1931
3	3	5	$\tau^{-4}/16$	$332 + 144\sqrt{5}$	0.07819
4	3	4	0	∞	0
3	3	6	$-\frac{1}{16}$	$52 + 24\sqrt{3}$	0.2064
4	3	5	$-\tau^{-1}/8$	$18 + 5\sqrt{2} + 14\sqrt{5} + 5\sqrt{10}$	0.2349
3	5	3	$-(2\sqrt{5}-3)/16$	$34 + 18\sqrt{5}$	0.2316
3	4	4	$-\frac{1}{8}$	$28 + 18\sqrt{2}$	0.2727
4	3	6	$-\frac{1}{8}$	$16 + 6\sqrt{2} + 8\sqrt{3} + 2\sqrt{6}$	0.3030
5	3	5	$-(5\sqrt{5}-7)/32$	$18 + 10\sqrt{5}$	0.3135
5	3	6	$-\tau^2/16$	$58 + 30\sqrt{3} - 18\sqrt{5} - 10\sqrt{15}$	0.3574
3	6	3	$-\frac{3}{16}$	$16 + 12\sqrt{3}$	0.3283
6	3	6	$-\frac{3}{16}$	$12 + 8\sqrt{3}$	0.3909
4	4	4	$-\frac{1}{4}$	$12 + 10\sqrt{2}$	0.3887

TABLE 3. TABLE OF HONEYCOMBS



p	r	$-\Gamma$	$\Sigma\Sigma c_{\mu\nu}$	2ϕ
3	3	0	∞	0
3	4	$(2\sqrt{2}+1)/16$	$12+16\sqrt{2}$	0.3382
3	5	$(5\sqrt{5}+1)/32$	$4+8\sqrt{5}$	0.4243
3	6	$(\sqrt{3}+2)/8$	$4+8\sqrt{3}$	0.4689
4	4	$\frac{7}{16}$	$8+8\sqrt{2}$	0.4512
4	5	$(\sqrt{2}+1)(\sqrt{2}+\sqrt{5})/16$	$10\sqrt{2}-2\sqrt{5}+10\sqrt{10}-26$	0.5060
4	6	$(\sqrt{2}+\sqrt{3})^2/16$	$64-48\sqrt{2}+44\sqrt{3}-24\sqrt{6}$	0.5371
5	5	$(7\sqrt{5}+5)/32$	$8\tau=4+4\sqrt{5}$	0.5490
5	6	$(2\sqrt{3}+2\sqrt{5}-1)\tau/16$	$4+20\sqrt{3}-12\sqrt{5}$	0.5741
6	6	$\frac{1}{4}$	$4+4\sqrt{3}$	0.5961

TABLE 4. TABLE OF HONEYCOMBS



p	q	$-\Gamma$	$\Sigma\Sigma c_{\mu\nu}$	2ϕ
3	3	$\frac{1}{4}$	$9+6\sqrt{2}$	0.4738
3	4	$(\sqrt{2}+1)^2/8$	$4+6\sqrt{2}$	0.5587
4	4	1	$4+4\sqrt{2}$	0.6330

According to these tables, the smallest 2ϕ (namely, 0.07819) occurs in case (i) with $p = q = 3$ and $r = 5$, when the points are the in-centres of the 14,400 tetrahedra



which can be packed together to fill spherical 3-space (Goursat 1889, p. 99; Coxeter 1948, pp. 153, 232). This means that the points are the vertices of the uniform spherical honeycomb



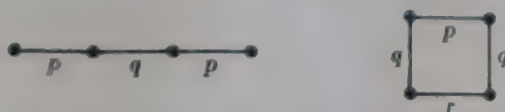
whose cells consist of

- 1200 hexagonal prisms,
- 720 decagonal prisms,
- 600 truncated octahedra,
- and 120 truncated icosidodecahedra.

(For the deduction of the cells from the graphical symbol, see Coxeter 1935, p. 329.) Since the spherical space-groups are all known, we can be quite sure that *this distribution of points has the finest possible mesh for spherical space*. In other words, the greatest possible number of points in a non-trivial homogeneous distribution in spherical space is equal to the order of the largest group, namely, 14,400.

But in hyperbolic space there remains the possibility of a close-grained distribution whose group is not generated by reflexions but is derived from a reflexion group

by adjoining one or more rotational symmetry operations of the fundamental region.
The tetrahedra



are symmetrical by a half-turn about the line joining the mid-points of a pair of opposite edges; and when $r = p$ the latter is symmetrical by three such half-turns belonging to a 4-group. After adjoining these half-turns, we obtain a new distribution of points consisting of the centres of a packing of equal spheres, two or four inside each tetrahedron. The spheres touch one another, and each touches three faces of the tetrahedron. The computation of their radius ϕ would be quite formidable, but we can obtain an adequate approximation by comparing this sphere-packing with that in which each tetrahedron contains only one sphere, whose radius ϕ' is given in the above tables. Since the same volume now contains 2 or 4 times as many spheres, about as densely packed as before, ϕ'/ϕ will be approximately $\sqrt[3]{2}$ when the tetrahedron contains two spheres, and $\sqrt[3]{4}$ when it contains four.

The validity and limitations of this use of $\sqrt[3]{2}$ can be assessed by considering the Euclidean tetrahedron

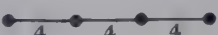


for which we can easily compute the exact radii of one or two spheres packed inside. The ratio is found to be

$$\phi'/\phi = 1 - \sqrt{\frac{1}{3}} + \sqrt{\frac{2}{3}} = 1.2391,$$

whereas $\sqrt[3]{2} = 1.2599$.

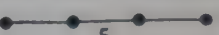
Similarly, the use of $\sqrt[3]{4}$ can be checked by means of the dissection of the hyperbolic

tetrahedron  into four 's, which shows that four spheres in

the former have each the same size as one in the latter. Calling the radii ϕ' and ϕ , we have (from tables 2 and 4)

$$\phi'/\phi = 6330/3887 = 1.628,$$

whereas $\sqrt[3]{4} = 1.5874$. (The discrepancy is due partly to the curvature of the space, but chiefly to the different densities of the two sphere-packings.)

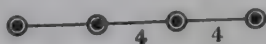
By fitting four spheres into  we can expect to reduce 2ϕ from 0.4512 to about $0.4512/\sqrt[3]{4} = 0.282$; but this is larger than some of the values already obtained. More significantly, by fitting two spheres into  we can expect to reduce 2ϕ from 0.2316 to about

$$0.2316/\sqrt[3]{2} = 0.184,$$

which is appreciably less than the in-radius of . In fact, looking over all the tables, we are led to conclude that 0.184 is the smallest possible spacing for the points of a non-trivial homogeneous distribution in hyperbolic space. But we cannot

assert this with the same absolute confidence as the corresponding result for spherical space, since we must admit the highly improbable possibility of a slightly better distribution given by a group whose reflexions and rotations are mixed in a more complicated manner.

Among the moderately close-grained hyperbolic honeycombs, the only one readily amenable to analytic treatment is



($2\phi = 0.2727$), which has the same symmetry group as the regular honeycomb $\{4, 4, 3\}$ of § 3. We again use co-ordinates (x_0, x_1, x_2, x_3) , with the absolute quadric

$$x_0^2 - x_1^2 - x_2^2 - x_3^2 = 0,$$

so that the symmetry group is generated by the permutations of x_1, x_2, x_3 , their changes of sign, and the Lorentz transformation

$$x'_i = x_i + (x_0 - x_1 - x_2 - x_3).$$

A typical vertex of the honeycomb is the in-centre of the tetrahedron bounded by the planes

$$x_1 = x_2, \quad x_2 = x_3, \quad x_3 = 0, \quad x_0 = x_1 + x_2 + x_3.$$

Solving the equations

$$\frac{x_1 - x_2}{\sqrt{2}} = \frac{x_2 - x_3}{\sqrt{2}} = x_3 = \frac{x_0 - x_1 - x_2 - x_3}{\sqrt{2}},$$

we find this to be the point $(4\sqrt{2} + 3, 2\sqrt{2} + 1, \sqrt{2} + 1, 1)$. One of the neighbouring vertices is $(4\sqrt{2} + 3, 2\sqrt{2} + 1, \sqrt{2} + 1, -1)$, whose distance 2ϕ is given by

$$\sinh^2 \phi = \frac{1}{x_0^2 - x_1^2 - x_2^2 - x_3^2} = \frac{1}{28 + 18\sqrt{2}},$$

in agreement with table 2 for the case $p = 3, q = r = 4$.

5. CONCLUSIONS

We saw, in § 3, that several 'homogeneous and almost isotropic' distributions of points can be found both in spherical and in hyperbolic space. The distance between neighbouring points is roughly an absolute unit. In § 4 we saw that by abandoning *local* isotropy we could only reduce the distance to about one-fifth of the absolute unit in the hyperbolic case and to about one-thirteenth in the spherical. However, the *order of magnitude* of the empirically determined ratio of the average distance between neighbouring nebulae (or clusters) to the radius of curvature of astronomical space appears to be not greater than about one part in a thousand (or a hundred). We conclude therefore that, given the empirical order of magnitude of the ratio of the distances between neighbouring nebulae (or clusters) to the 'radius of the universe', *the observed irregularity in the distribution of nebulae (and clusters) is not surprising, since some irregularity is geometrically inevitable unless, of course, the nebulae can be represented by a static distribution of particles in space of zero*

curvature. As usual, we neglect the proper motions of individual nebulae (and clusters), which we assume are small. Consequently, if our honeycombs had any astronomical meaning, their vertices could only represent super-clusters for whose existence there is at present no evidence.

A further problem which remains to be investigated is to discover why the distances between neighbouring nebulae are so small, i.e. why space is so closely packed with galaxies. Under certain assumptions about the average density of matter in space, this question may be regarded as roughly tantamount to the following: Why do the great nebulae contain the equivalent of about 10^{11} stars of the mass of the sun?

REFERENCES

- Coxeter, H. S. M. 1935 *Proc. Lond. Math. Soc.* (2), **38**, 327–339.
 Coxeter, H. S. M. 1939 *Trans. Amer. Math. Soc.* **45**, 73–150.
 Coxeter, H. S. M. 1940 *Math. Z.* **46**, 380–407.
 Coxeter, H. S. M. 1942 *Non-Euclidean geometry*. Toronto University Press.
 Coxeter, H. S. M. 1948 *Regular polytopes*. London: Methuen.
 Fricke, R. & Klein, F. 1897 *Vorlesungen über die Theorie der Automorphen Funktionen*, **1**. Leipzig: Teubner.
 Goursat, E. 1889 *Ann. Sci. École Norm. Sup.* (3), **6**, 9–102.
 Klein, F. 1879 *Ges. math. Abh.* **3** (Berlin, 1923), 126, 131.
 McCrea, W. H. 1931 *Proc. Edinb. Math. Soc.* **2**, 158.
 Milne, E. A. 1934 *Quart. J. Math.* **5**, 30.
 Milne, E. A. 1935 *Relativity, gravitation and world structure*. Oxford University Press.
 Milne, E. A. 1937 *Proc. Roy. Soc. A*, **158**, 324–348; **159**, 171–191.
 Omer, G. C. Junr. 1949 *Astrophys. J.* **109**, 164.
 Robertson, H. P. 1935 *Astrophys. J.* **82**, 284–301.
 Schild, A. 1949 *Canad. J. Math.* **1**, 29–47.
 Schlegel, V. 1883 *Nova Acta Leop. Carol.* **44**, 343–459.
 Sommerfeld, A. 1909 *Verh. dtsch. Phys. Ges.* **11**, 577.
 Sommerville, D. M. Y. 1924 *R.C. Circ. mat. Palermo*, **48**, 17.
 Sommerville, D. M. Y. 1929 *An introduction to the geometry of n dimensions*. London: Methuen.
 Stringham, W. I. 1880 *Amer. J. Math.* **3**, 1–14.
 Threlfall, W. & Seifert, H. 1931 *Math. Ann.* **104**, 1–70.
 Tolman, R. C. 1934 *Proc. Nat. Acad. Sci., Wash.*, **20**, 169.
 Varićak, V. 1910 *Phys. Z.* **11**, 93, 287, 586.
 Walker, A. G. 1937 *Proc. Lond. Math. Soc.* (2), **42**, 90–127.
 Whitrow, G. J. 1935 *Quart. J. Math.* (Oxford), **6**, 249–260.
 Whitrow, G. J. 1936 *Z. Astrophys.* **12**, 47–55.
 Whitrow, G. J. 1937 *Z. Astrophys.* **13**, 113–125.
 Whitrow, G. J. 1939 *Quart. J. Math.* **10**, 313–319.

WILKINS LECTURE

Robert Hooke

BY E. N. DA C. ANDRADE, F.R.S.

(Delivered 15 December 1949—Received 24 January 1950)

[Plates 12 to 14]

Science in England in the latter part of the seventeenth century is overshadowed by the mighty name of Newton, who has justly received the praises of all the great natural philosophers who came after him. In that springtime of science there were, however, in England a number of other men of genius who carried out work of prime importance—Robert Boyle; John Wallis and Isaac Barrow; Flamsteed and Halley; Willughby and Ray; Sydenham and Glisson; and Robert Hooke. Of these Robert Hooke has good claims to be considered the greatest. Probably the most inventive man who ever lived, and one of the ablest experimenters, he had a most acute mind and made astonishingly correct conjectures, based on reason, in all branches of physics. Physics, however, was far from being his only field: he is the founder of scientific meteorology; as an astronomer he has observations of great significance to his credit; he did fundamental work on combustion and respiration; he was one of the founders of modern geology. He has, moreover, a particular claim to the attention and respect of our Society, for from 1662 to 1677 he held the office of Curator and from 1677 to 1682 he was one of our Secretaries. He was always indefatigable in his services to the Society, and for a period he produced new experiments or discoveries at practically every meeting. Most writers who have really studied his work have given Hooke enthusiastic praise, yet, on account of certain difficulties of character—difficulties which he was not the only one to possess—his name does not seem to be honoured as it should be among men of science in general. No one has ever devoted a book to his life and achievements,* but he has been made the subject of casual and ill-considered criticism. It therefore seemed to me that it would be altogether fitting that I should attempt to recall to you something about this extraordinary man; about his services to science and his services to our Society.

Robert Hooke was born at Freshwater, in the Isle of Wight, on 18 July 1635, his father being curate of the parish.† Aubrey says that his father was one of the family of the Hookes of Hooke in Hants. Hooke was thus seven years older than

* The standard account is the *Life*, only some twenty-eight pages long, which Richard Waller prefixed to the *Posthumous Works* (1705). This is cited in further references as *Waller*, and the page given in Roman numerals, as used in the *Life*. The *Life* in Ward's *Lives of the Gresham Professors* (1740) is largely taken from this. Dr J. R. Morgan wrote a praiseworthy thesis on *The Contributions of Robert Hooke to the Physical Sciences*, in 1930, but this has never been published.

† No trace remains of the house of his birth, although it is traditionally believed to have been on the south side of a hill known to-day as Hook Hill.

Newton, a fact which probably had some influence on the relations between the two men. Like Newton he was a weakly child, but whereas Newton grew up strong and straight, Hooke was never physically sound. We have a description of him from Richard Waller, who was Secretary of our Society from 1687 to 1709 (as well as again at a later period) and must have known him well. He tells us that as to his person he was but despicable, which recalls Samuel Pepys' entry for 15 February 1664/5, where, after telling us that he was that day admitted to the Royal Society 'by signing a book and being taken by the hand of the President, my Lord Brouncker, and some words of admittance said to me', he adds, 'Above all, Mr Boyle was at the meeting, and above him Mr Hooke, who is the most, and promises the least, of any man in the world that ever I saw.' Hooke was very bent and crooked, but told Waller that he was straight until he was about 16, when he grew awry by working at the lathe. Hooke, who left some notes about his early life, also said that as a boy he was very sprightly and active in running and leaping 'tho' very weak as to any robust Exercise'. 'He went', says Waller, 'stooping and very fast having but a light Body to carry and a great deal of Spirits and Activity, especially in his Youth.' He was also, one gathers, meanly ugly, very pale and lean: 'His Eyes grey and full, with a sharp ingenious Look whilst younger; his Nose but thin, of moderate height and length; his Mouth meanly wide and upper Lip thin; his Chin sharp and Forehead large... He wore his own Hair of a dark Brown colour, very long and hanging neglected over his Face, uncut and lank.' Aubrey, who seems to have been his close friend and most anxious to speak well of him—he says that he was a person 'of great suavity and goodness'—also records that he was something crooked, that his head was large but the lower part of his face little and that his grey eyes were 'full and popping'. I think it important that you should know something of his appearance and great physical disabilities, and I quote so fully from the descriptions of those that knew him because I can say with some confidence that there is no known portrait of any kind of him, although in his diary* he seems to suggest that one Bonus (usually spelt Bownest), a known artist, drew his picture. It is one of my ambitions to find that picture.

The infirm boy, like Newton, showed a great taste for making mechanical models and for drawing. It is recorded that he was, as a boy, for some time an apprentice to Sir Peter Lely, the artist, but the smell of the oil colours provoked the headaches which worried him through his life. Whether he was actually apprenticed to Lely or no, he was certainly a competent draughtsman, as the drawing reproduced in plate 12 and his architectural drawings bear witness. We get a glimpse of him at Westminster, where he acquired a passable knowledge of Greek and Latin and studied mathematics. At Westminster he lived with the famous headmaster Dr Busby, and it is worth noting that not only did Busby, as far as one can judge, like and encourage him but that later in life he and Busby were on intimate terms and often dined together. Hooke had some firm friends.

In 1653, at the age of eighteen, he went to Oxford, with no private means. He had a chorister's place at Christ Church, 'which', says Aubrey, 'was pretty good maintenance'. The position of chorister was, however, a lowly one; he was also servitor

* Diary of 1672 to 1680, edited by the care of our Mr H. W. Robinson and Mr W. Adams.

to a certain Mr Goodman. This subordinate social position of Hooke's is to be remembered when we come to consider his status in the Royal Society.* It is generally held that Dr Thomas Willis, of Christ Church, who had equipped a laboratory, soon employed Hooke as assistant and recommended him to Robert Boyle. In any case it is as Boyle's assistant that Hooke's scientific career begins.

Oxford was at that time the great centre of the new experimental philosophy. From about 1648 onwards a brilliant band of men, including John Wallis, Seth Ward, William Petty, Thomas Willis and John Wilkins, used to meet to discuss the new method of experiment; after Petty and Wilkins had left Oxford, the meeting place was at Robert Boyle's house. Just when Hooke went to help Boyle is not known; if the young man whom John Wilkins recommended to Boyle in a letter was actually Hooke, the year was 1653.† It was about 1655, however, that he began to mix freely with the great men who were enthusiastically experimenting and that he began to be known for his genius in mechanical devices.

In 1658 or 1659 he made for Robert Boyle the first air pump used by Boyle in his experiment, which was also the first air pump made in England. It was not, of course, the first air pump, for in 1657 Schott had published an account of Guericke's pump, but it was a great improvement on this.‡ Hooke says: 'I contriv'd and perfected the Air Pump for Mr Boyle, having first seen a Contrivance for that purpose made for the same honorable Person by Mr Gratorix, which was too gross to perform any great matter.'§ Boyle, in his *New Experiments Physico-Mechanicall* (1660), acknowledges that Hooke 'fitted him with' a pump, after 'Mr. G.' had failed. This pump, although crude in design compared to the mechanical contrivances of Hooke's maturity, acted so well that the workmanship must have been excellent. About the time of the making of the pump, that is, 1658 or 1659, Hooke was much occupied by projects for flying. It is typical of Hooke's perspicacity that he abandoned these projects because he decided, as a result of trials and calculations, 'that the Muscles of a Man's Body were not sufficient to do anything considerable of that kind'.

From now on we are to be confronted with the difficulty of coping with the stream of inventions, notions, brilliant suggestions, accurate observations, daring speculations and prophetic conjectures that poured from Hooke's fertile brain and contriving hands. It will be impossible even to mention them all; to classify them will be difficult; in many cases, in view of the scanty record, it will be hard to decide exactly what was done. Practically everything, however, will bear witness to a truly extraordinary inventiveness and a truly modern outlook. Sometimes Hooke is wrong, but he is wrong in a strictly scientific and not a medieval way. Very often the ideas which he tumbled out in such profusion were taken by others; sometimes his findings were reached quite independently by others, which Hooke

* In some ways, as, for instance, in his lowly position at Oxford; in his leaving without a degree, which was subsequently granted; in the doctor's degree conferred late in life; in his physical disabilities, Hooke makes us think of Samuel Johnson.

† See L. T. More, *Life of Boyle*, pp. 79, 80.

‡ Guericke's own account, describing a somewhat more elaborate pump, first appeared in 1672.

§ Waller, p. iii.

found hard to believe. At every stage we are witnessing the workings of a mind so active, so fertile in expedients, so interrupted at every hour, at every endeavour, by the inrush of new concepts, new projects, that it is hard to disentangle his doings. Newton said that he made his discoveries by keeping the subject constantly before him and waiting until the first dawnings opened little by little into the full light. This Hooke was quite unable to do; he totally lacked Newton's powers of concentration. His mind was restless, continually disturbed by fresh ideas, but they were nearly all good and many were of first importance.

In 1661, while he was still with Boyle, appeared a little book of fifty small octavo pages, *An Attempt for the Explication of the Phenomena Observable in an Experiment Published by the Honorable Robert Boyle*, which dealt with the rise of liquids in capillary tubes and with surface-tension phenomena in general, for Hooke recognized that the spherical form of drops, the rise of water at the edge of a vessel and the rise in fine tubes were all aspects of the same general behaviour. His explanations were wrong: he put everything down to the pressure of one fluid on the other; thus the spherical drop was due to pressure of the air. In particular, the rise in the capillary tube was due to the pressure of the air being less within the fine tube than in the room at large. In this connexion he carried out experiments on the difficulty of forcing air through fine tubes—on the viscosity of air, as we should say. If you smile at the notion that the pressure within a fine open tube may be less than outside, remember that Newton accepted it.* Hooke's little tract is full of acute suggestions; he says that the capillary rise explains the rise of oil in the wick of a lamp and possibly the ascent of sap in trees, and sets down many accurate observations on the behaviour of small floating bodies, due to surface tension (which, needless to say, he does not so name). Clearly it is not possible here to discuss this minor work of Hooke's, but I cannot resist quoting the words in which he gives his method of work: 'For I neither conclude from one single Experiment, nor are the Experiments I make use of, all made upon one Subject: Nor wrest I any Experiment to make it *quadrare* [square] with any pre-conceiv'd Notion. But on the contrary, I endeavour to be conversant in all kind of Experiments, and all and every one of those Trials, I make the standards (as I may say) or Touchstones by which I try my former Notions. . . .' Thus the young man of twenty-six set down his scientific creed in an age when the experimental method was still making its way.

Now in these early days Hooke made an invention of capital importance, the history of which was most unfortunate for him, namely, the invention of the balance spring for watches. Until this time the portable spring-driven watch was governed by a swinging bar, which rebounded to and fro as the pallets struck the teeth of the escape wheel. This is clearly a very rough control; the error in the best of these watches was about a quarter of an hour a day.

* In 1675, in *An Hypothesis explaining the Properties of Light, discoursed of in my several Papers*, Newton says: 'as the air can pervade the bores of small glass pipes, but yet not so easily as if they were wider; and therefore stands at a greater degree of rarity than in the free aerial spaces, and at so much a greater degree of rarity as the pipe is smaller, as is known by the rising of water in such pipes to a much greater height than the surface of the stagnating water, into which they are dipped.'

Hooke, whose interest in astronomy was awakened by Seth Ward, had been experimenting with pendulum clocks as accurate time-keepers and, according to his account, conceived the idea of springs as a control for watches, with the longitude problem in his mind. He asked for Boyle's help to protect his invention, Boyle invoked Lord Brouncker and Sir Robert Moray, and an agreement was drawn up somewhere about 1660 between Boyle, Brouncker, Moray and Hooke for the marketing and protecting of what was, in Hooke's words, 'a pocket watch, accomodated with a Spring, apply'd to the arbor of the Balance to regulate the motions thereof'. According to Hooke's subsequent account* the negotiations were broken off because he would not agree to a clause in the document drawn up by Moray stating that if, after Hooke had disclosed his invention, any other persons should find a way of improving the principles, they should have the benefit thereof during the time of the patent. Hooke contended that it was easy to vary his principles and perhaps to effect improvements (it being, as he said, *facile inventis addere*) and that this clause might deprive him of the benefit of his invention. The existence of this document is well attested. I do not think that an agreement of this kind would have been drawn up unless a going watch had been produced. Hooke was a young man without social position, protection or influence; the other three were not business men ignorant of science and anxious to make a lucky speculation, but critical spirits apprised of the science of their time. Robert Boyle is known to all as the author of the *Sceptical Chymist*; Brouncker was a first-rate mathematician and first President of the Royal Society; and Moray, if not of this calibre, had extensive scientific knowledge† and was a man of the world. Incidentally he held office as President unofficially at the time when the Society was in its primitive stage.

If I insist upon the significance of this document it is because the subsequent history of the spring-governed balance wheel is involved. Hooke says that he showed the watch described in the words already quoted to the three men interested in the patent, but there is no other record. In *The History of the Royal Society* by Thomas Sprat, dated 1667, is a list of instruments invented by the Fellows, given without the names of the originators, which contains several known inventions of Hooke's; in this is recorded 'Several new kinds of *Pendulum Watches* for the Pocket, wherein the motion is regulated by Springs or Weights, or Loadstones, or Flies moving very exactly regular.' The use of loadstones and 'watch flies' is known for certain to have been put forward by Hooke. A letter dated 30 September 1665, from Sir Robert Moray to Oldenburg, very definitely infers that Hooke had already applied a spring to the arbor of the balance.‡ I consider it practically certain that Hooke had invented the modern spring-controlled balance wheel by 1665, and that earlier he had a watch controlled by a straight spring attached to the balance wheel. I think it very probable that working watches had been made under his supervision—there is a tradition that he presented one to Wilkins, after whom this lecture is

* In *A Description of Helioscopes and some other Instruments*, 1676.

† 'Authorities concur in assigning to him an extensive knowledge of natural philosophy and mathematics', Weld, *History of the Royal Society*.

‡ This letter, from which Waller quotes on p. vi of the *Life*, is still in the possession of the Royal Society (M.1.13).

named, about 1661. He frequently affirmed that he had made a watch more exact than any pendulum clock, but does not seem to have produced it when requested by the Society to do so. The whole matter was brought to a head when Huygens, in 1674, brought out a watch with a balance wheel governed by a spiral spring. His application for a French patent failed because Hautefeuille claimed that he had already used springs in this way. Huygens's invention revived Hooke's interest in the balance wheel, he busied himself with Tompion making watches, and in 1675 he presented to Charles II a spring-controlled watch, inscribed *Rob. Hooke invenit 1658 T. Tompion fecit 1675*, which, of course, is no proof that he invented it in 1658. He referred his claim to priority to the Society, which did not support him, but seemed to favour Huygens. In this connexion it must be remembered that Oldenburg, the Secretary from 1663 to 1677, was an enemy of Hooke's, who undoubtedly made mischief between Hooke and Newton, and that further he had a financial interest in Hooke's discomfiture, for Huygens had assigned to Oldenburg his English patent rights, and Oldenburg tried in vain to get an English patent granted. But even Oldenburg, writing in 1675, admits 'Tis certain then, that the Describer of the Helioscope' (Hooke) 'some years ago caus'd to be actually made some Watches of this kind', adding, 'but it is as certain that none of these Watches succeeded', which he cannot have known. Hooke gave a circumstantial account of his invention in *A Description of Helioscopes*, published in 1676. The episode is of importance, since Hooke felt his lack of support from the Royal Society very deeply, and especially the part which Oldenburg played in it.

If I have devoted some time to this question of the balance wheel it is not so much because it is of capital importance—for Hooke made many discoveries of capital importance—but because it is an example of the difficulty of settling precisely the history of Hooke's inventions. He would come upon some brilliant and original mechanical conception, take it a certain distance, generally far enough to show that it would work, and then, stopping short of the final stage and precise publication that would convince the world of his priority, take up some other brilliant notion that had come into his head. He would then, naturally, feel aggrieved when someone else produced an elaboration or great extension of his work, generally without any acknowledgement. It is clearly impossible to discuss, however briefly, the history of each of his inventions. What I now propose is to describe briefly to you the course of his life and then to try to estimate his scientific achievement.

I have already referred to Hooke's life at Oxford as paid assistant to Robert Boyle. The Royal Society, having received its Charter in 1662, set about the appointment of a Curator whose duty was 'to furnish the Society every day they mett, with three or four considerable experiments, expecting no recompense till the Society gett a stock enabling them to give it'. Robert Hooke was appointed to the post. From this time onwards he devoted much of his time and much of his invention to the Society's interest; it is scarcely an exaggeration to say that without his experimental demonstrations, which over long periods took place weekly, the Royal Society would have died, as did the Accademia del Cimento, or declined into scientific insignificance, as did the Académie des Sciences on the death of Colbert in 1683.

To turn the pages of Birch's *History of the Royal Society*, which is a record, sometimes a very detailed record, of the meetings, week by week, from 1660 to 1687, is to convince oneself that Hooke's experiments and theoretical ideas, and the discussions which they provoked, were the main agent that held the Society together.

In 1664 it was voted that he should be given a regular salary, with the help of Sir John Cutler's promise of £50 a year. The record of Hooke's payments is, like everything else to do with him, scattered and complicated; in general it may be said that the payments were poor and irregular. Typical is that Cutler never paid what he promised, so that Hooke had to enter on a Chancery suit against Cutler's heirs to obtain his dues. In 1696, about thirty years after Cutler's first promise, the suit was determined in Hooke's favour, an event that gave him the greatest satisfaction. On one occasion, the Society proposed to pay Hooke in copies of Willughby's *Historia Piscium*, a work printed at the expense of the Society of which a large number of copies were left unsold.* Hooke apparently lived in the early days of the Society by consulting work, as we should call it now, done for instrument makers and technologists of all kinds. After the Great Fire he had lucrative employment in connexion with the rebuilding schemes.

In 1665 appeared his great work *Micrographia*, inscribed to Charles II in a short preface that displays Hooke as a writer of polished and beautiful prose. This book, as the title would suggest, contains a record of a large number of microscopical observations, made with the assistance of a compound microscope, which, judging by the plates of objects seen through it, must have been an excellent instrument. An original Hooke microscope, of the type described in the *Micrographia*, is shown in plate 13.

The plates in the *Micrographia* are beautiful in themselves, but also record a number of fundamental discoveries, to some of which I shall refer later. Sachs, the historian of botany, puts Hooke with Malpighi, Grew and Leeuwenhoek as 'endeavouring by earnest reflection to apply the powers of the mind to the objects seen with the assisted eye, to clear up the true nature of the microscopic objects, and to explain the secrets of their constitution'. The figures of the gnat, the flea and the louse were long famous. But microscopic pictures and their discussion form but a small part of the book. In it we find important theoretical discussions of the nature of light and heat, on which I shall animadvert later; further discussion of capillarity on the lines of his earlier tract; experiments on the thermal expansion of solids and liquids; shrewd speculations on tempering of metals; observations on crystal structure; astronomical discussions, including attempts to form artificially craters like those of the moon; and accounts of the magnitudes of stars, in which occurs the statement that more powerful telescopes would discover fresh stars: 'I am apt to think, that with longer Glasses, or such as would bear a bigger aperture, there might be discovered multitudes of other small Stars, yet inconspicuous. And, indeed, for the discovery of small Stars, the bigger the aperture be, the better

* It was the printing of this book that so exhausted the Society's funds that there was nothing available to pay for the printing of Newton's *Principia*; in consequence, Halley undertook to print it at his own charge. For payments in 'books of fishes' see Weld, *History of the Royal Society*, I, 310.

adapted is the Glass.' Further, we must note that the book contains a very full discussion of the colours of thin plates, such as flakes of mica, air films between glasses, and bubbles not only of soapy water, but of rosin and several other substances. These observations were the cause of subsequent dispute with Newton.

The *Micrographia* gained Hooke considerable fame at home and abroad. Pepys was delighted with it; it received a lengthy review in the *Journal des Sçavants*, with folding copies of two of the plates. Sturm in his *Collegium Experimentale*, published in 1676, praises it in the highest terms, saying that he has seen an English microscope and that whoever does not admit that England bears the palm for this kind of instrument is either envious or has not seen Hooke's *Micrographia*. He gives a folding plate of a Hooke microscope.

In 1665 the Great Plague caused the discontinuance of the weekly meetings of the Royal Society, and it appears that Hooke went with other Fellows to Banstead Downs in Surrey, where many experiments were made. After the Great Fire in 1666 Hooke's abilities as an architect and surveyor obtained for him the lucrative post of City Surveyor, which brought him into close contact with Sir Christopher Wren—in fact, it is not too much to say that Hooke was Wren's chief assistant in all matters concerning the rebuilding.* In Hooke's *Diary* for 1672 to 1680, Wren is mentioned on nearly every page. Very frequent, too, is the entry of a 'view' for 10 shillings, which refers to a survey of a house foundation. It is meet that I should here refer to Hooke's eminence as an architect. Among other buildings designed and built by him were Bedlam Hospital, the Haberdashers' Alms-Houses, and Montague House, praised by Evelyn: 'On the whole it is a fine palace, built after the French pavilion-way by Mr Hooke.' It stood on the site of the British Museum, and was burnt down in 1686. Incidentally, Hooke produced a method of making bricks, with less clay and more speed than had been done before, at the Royal Society in 1667, but I can find no more of it. He even prepared a model for rebuilding the city, which appears to have represented a plan by which all the chief streets should be laid out on a rectangular or 'gridiron' pattern, as in Mannheim and New York. It is characteristic that, while the Great Fire ceased on 12 September 1666, this model was laid before the Royal Society on 19 September. The Society was well pleased with it.

It was in 1672 that the unfortunate disputes with Newton took place concerning the experiments on the decomposition of white light. At the time of the preparation of the *Principia* they flared up again. I touched upon them when I had the honour of lecturing before you on Newton.† I do not, I think, underrate the difficulties that might easily occur in intercourse between two such men as Hooke, touchy and anxious for his priority, and Newton, secretive, self-centred, assured and desperately averse from all controversy. Nevertheless, each man found much to respect in the

* '...disposed him to take to his Assistance Mr. Robert Hooke, Professor of Geometry in Gresham College, to whom he assigned chiefly the Business of measuring, adjusting and setting out the Ground of the private Street-houses to the several Proprietors...' *Parentalia*, p. 263. There is evidence for the opinion that Hooke was responsible for the design of the Monument.

† *Newton Tercentenary Celebrations* of the Royal Society, 1942 and 1946.

other—Newton in certain letters refers to Hooke in a very complimentary manner and was, from his copious notes which still exist, a close student of the *Micrographia*. Hooke could not but admire Newton's prodigious experimental skill and knew, I am sure, well enough that he could never rival Newton as a mathematician or in quantitative theory. Hooke, furthermore, did not long bear rancour if left unprovoked; I have already pointed out how aggrieved he felt over Huygens's watch, yet later on he made very complimentary references to Huygens's work.

The trouble between the two men began with the publication, in February 1671/2, of Newton's first paper on the decomposition of white light by the prism. Newton was, at the time, unknown in London except for the reflecting telescope which he sent to the Royal Society in 1671. Hooke, the senior man, already firmly established as the author of the *Micrographia*, was concerned to defend his theory of colour and criticized Newton's interpretation of his experiments, although he admired the experiments. I have no doubt that Oldenburg, who hated Hooke—in my opinion as the man who dominated the Royal Society meetings where he wished to shine as the administrator on whom all depended—did all he could to make trouble between the two men. Hooke wrote a very civil letter to Newton in which he spoke of 'two hard-to-yeild contenders put together by the ears by other's hands and incentives' and Newton wrote to Oldenburg 'Pray present my service to Mr Hooke, for I suppose there is nothing but misapprehension in what lately happened', but it is unlikely that Oldenburg ever conveyed the compliment; he was concerned to keep up the dispute. If only John Locke had been the intermediary!

All the time that he was working with Wren on the rebuilding of the city Hooke was actually engaged in experiment, although not quite such a regular attender at the Society meetings as before, which can be understood. The middle seventies saw him, however, particularly active scientifically. In 1674 he published *An Attempt to Prove the Motion of the Earth* with a very courteous dedication to Sir John Cutler, who had promised him emoluments which he never paid, and his *Animadversions on the First Part of the Machina Coelestis*, both works which I shall have occasion to consider later. In 1676 we have his *Description of Helioscopes*, with the motto *Sic vos non vobis*. *Lampas* and *Cometa* followed in successive years and in the same year as *Cometa*, 1678, appeared *Lectures de Potentia Restitutiva or of Spring*, which contains Hooke's law, the sole achievement of this fertile genius known to most young men of science. This tract contains a record of several other prime discoveries. All these were bound up together in 1679 as a book with the title *Lectiones Cutlerianae*, a further compliment to the bilking Cutler, who thus achieved immortality without cost to himself.

In 1677 Oldenburg died and Hooke was appointed to the Secretaryship of the Royal Society, which he gave up in 1682. In 1679, after the *Philosophical Transactions* had ceased to appear, Hooke, at the request of Council, undertook to publish a kind of scientific sheet mainly concerned with what was being done abroad. This publication, known as the *Philosophical Collections*, ran from 1679 to 1682. Hooke then handed over the Secretaryship to Robert Plot and the *Philosophical Transactions* again appeared. There is no reference to the *Philosophical Collections* in our *Record*—Hooke's usual fate.

It was as Secretary that, in 1679, Hooke approached Newton to ask for a scientific contribution to the Society in a letter which was not only courteous but showed a complete lack of any ill feeling. Newton replied in a letter that I have quoted before, saying that he had lost interest in science.* In this letter Newton made a slip about a falling body on a moving earth, and Hooke corrected him, which irritated Newton. In a letter of 6 January 1680, Hooke definitely suggested the inverse square law to Newton, who, it is at the present day plain, was already in possession of it, but Hooke could not know that. Hooke was likewise perfectly clear on the basic principles on which the problem of planetary motion was to be solved, as I shall show when I come to discuss Hooke's scientific performances, but lacked the mathematical power to prove Kepler's laws. To Hooke's very courteous letter of 17 January 1680, which once more put forward the inverse square law, and asked for Newton's 'thoughts', Newton returned no answer. It is understandable, then, that Hooke, who could not have known that Newton already had the outline of the planetary theory in his head when he was at Woolsthorpe as a young man, was irritated when he learnt, six years after this letter, that Newton was going to publish the solution of the problem of planetary motion, on the lines which he had put forward, without any acknowledgement to him. All that he wanted, according to Halley's letter to Newton, was some mention in the preface, and, in support of his claim to this, at the least it seems quite possible that Hooke's correspondence set Newton's mind on the matter again. A civil word would have cost Newton nothing. He replied to Halley in two letters bitterly attacking Hooke and made no mention of him either in the *Principia* or in the *Opticks* (the publication of which he delayed until after Hooke's death), although it is quite certain from his written notes that he had carefully studied the *Micrographia*. Many of us will wish, with More, the biographer of Newton '... that he (Newton), in the full plenitude of his fame, could have shown more tolerance and a greater sympathy for that brave mind and spirit, housed in a suffering body'.

After giving up the Secretaryship Hooke continued to carry out experiments and to address the Society on all possible subjects, including the nature of memory and the notion of time, as well as the nature of comets. Everything that remains of his is full of interest and astonishingly modern in spirit. From about 1681 onwards it appears that he seldom left any full account of his lectures which could be entered in the Society's register, but intended later to write them up himself and publish, which he never did. That we have, in his *Posthumous Works*, a printed account of so much that he did which was never published during his life is due to Richard Waller, who brought out the volume with great care, preferring to print the manuscript just as the author left it rather than to amend or abbreviate. The book is dedicated to Newton, as President of the Royal Society! A further small volume containing papers by Hooke was brought out by Derham in 1726.

In 1687 he was much distressed by the death of his niece, Grace Hooke, and apparently became from then on 'more unactive, melancholy and cynical'. Nevertheless, he continued to contrive; for instance, in 1691 he was busy with instruments for sounding the sea and he built Aske's Hospital (Haberdashers' Alms-House).

* *Newton Tercentenary Celebrations, 1942*, p. 10.

He continued to lecture intermittently, when he was well enough, until 1696, when he seems to have become more or less a permanent invalid. The next year his case appeared hopeless, and both he and his friends thought that he would die forthwith. However, he continued in a miserable state—I will spare you such details as are recorded—until, in March 1703, he died in a condition of complete exhaustion. At his funeral all the Fellows then in town attended, ‘paying’, says the recorder, ‘the Respect due to his extraordinary Merit’. He was, it is set down, decently and handsomely interred in St Helen’s Church, Bishopsgate, where Sir Thomas Gresham also lies buried. The site of his grave is unmarked and unrecorded.

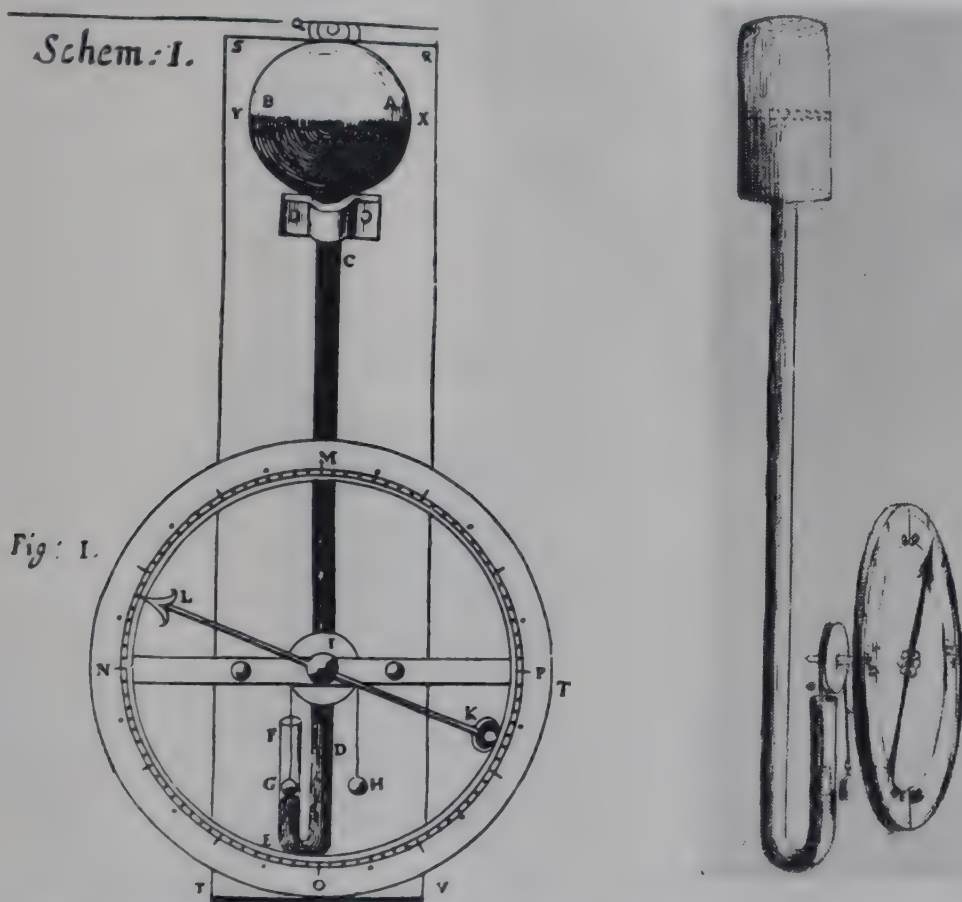


FIGURE 1. Hooke's wheel barometer: on the left is the diagram as given in the *Micrographia*; on the right Hooke's original sketch, in the possession of the Royal Society.

Let us now turn to Hooke's scientific achievement and consider first what Hooke did for the scientific instrument. It is in this field that he undoubtedly excels any other figure in the history of science. First of all, he invented all the meteorological instruments. His wheel barometer, shown in figure 1, is familiar to everybody; such barometers are still in use. Practically the last thing on which he busied himself was his marine barometer. As regards the thermometer, he proposed and used the freezing-point of water, ‘common distilled water, that is so cold that it just begins to freeze and shoot into flakes’, as the standard zero, the lower fixed point, just as it is used to-day. He had no upper fixed point, but avoided the necessity for one by a standard method of calibration. He prepared a cylindrical vessel, with a tube of one-tenth its diameter protruding from it, and marked on this tube graduations

representing one-thousandth of the volume of the vessel (figure 2). Using the same liquid in this as in his thermometer, he could clearly calibrate his thermometer in divisions representing an expansion of one-thousandth, and subdivide these. It was a properly standardized instrument that he made. His wind gauge was a plate pivoted so as to be blown aside by the wind, the strength of which was measured by the angular deflexion (figure 3). Instruments of this type were issued in the first

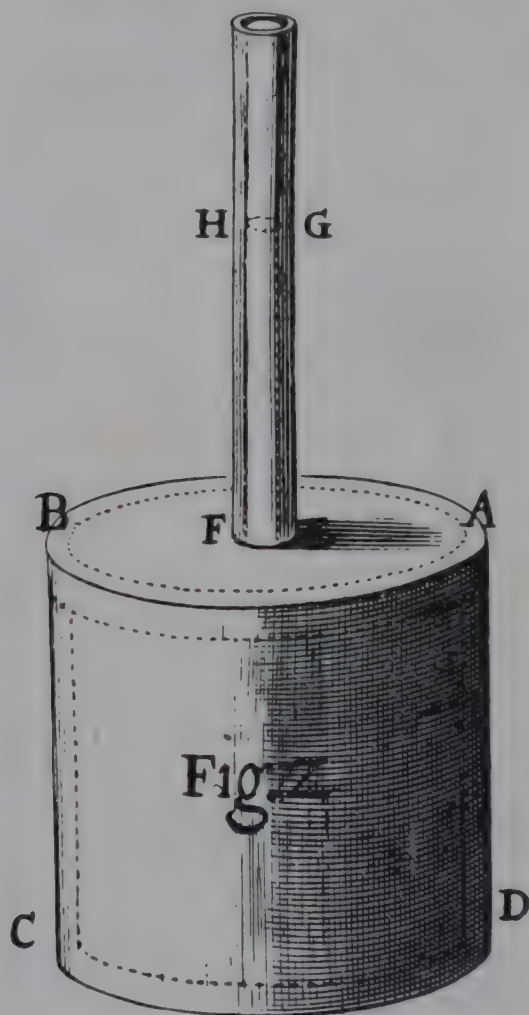


FIGURE 2. Hooke's vessel for use in calibrating thermometers. (From the *Micrographia*.)

world war. For measuring humidity, 'the dryness or moistness of the air', he employed the beard of a wild oat, which is naturally bent near the middle; the upper part rotates when the beard is moistened, owing to a natural twist in the structure of the lower part (figure 4). This beard he arranged so as to turn an index, which gave him his hygrometer. He also constructed a self-measuring rain gauge. He thus had at his disposal all the usual meteorological instruments, with the exception of a sunshine recorder. Further, following, as he freely acknowledged, the first design of Wren, he made an instrument which actually recorded the meteorological quantities. In this 'weather clock', which occupied him many years in the making, the instruments recorded every quarter of an hour on a paper strip by means of punches, the paper being carried on a rotating cylinder driven by a strong pendulum clock. According to the description, the barometer, the thermometer, the hygroscope, the rain bucket, and the wind vane all registered, the wind vane, of the

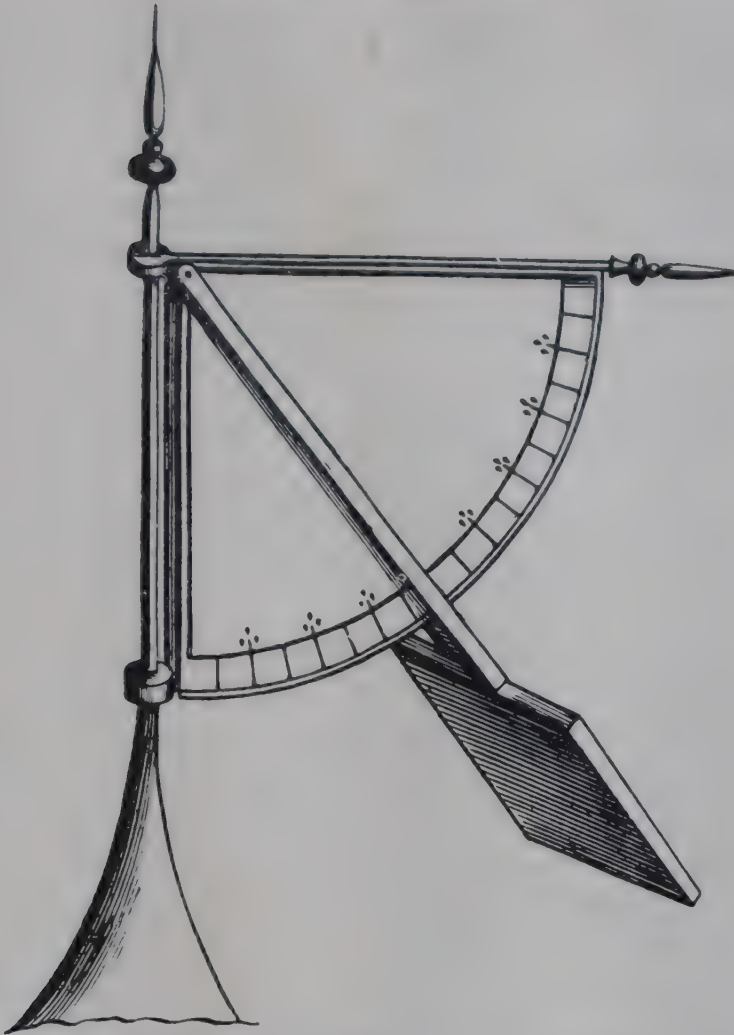


FIGURE 3. Hooke's wind gauge.

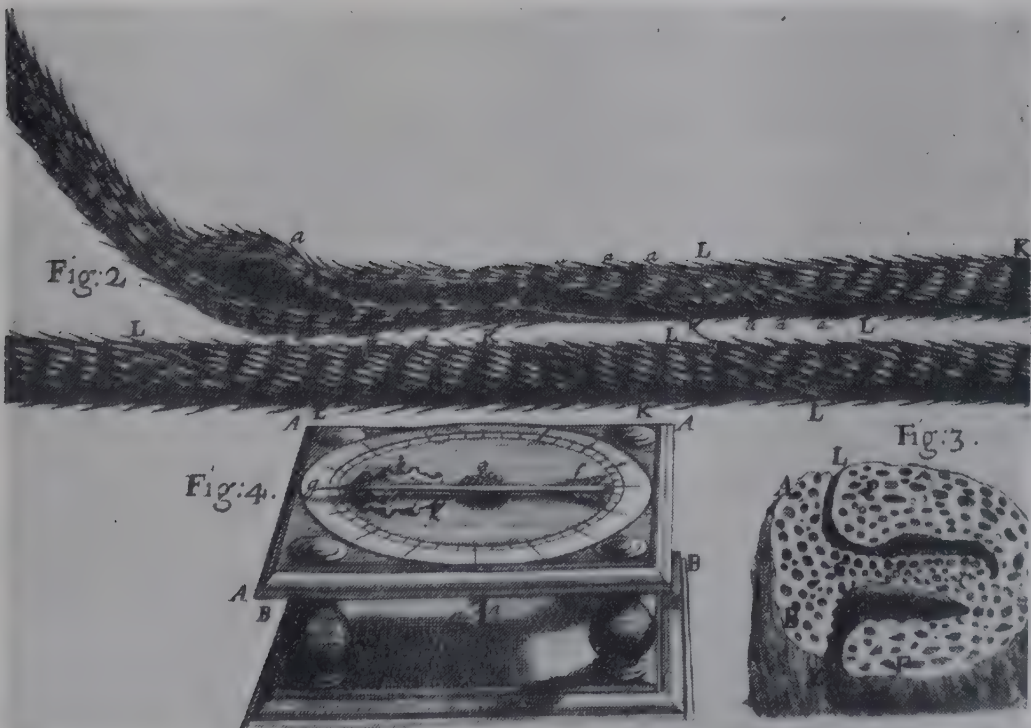


FIGURE 4. Hooke's hygrometer. In 'Fig: 2' is shown the 'wreathed', or twisted, beard of the wild oat, in two parts: in 'Fig: 3' a cross-section of the beard; in 'Fig: 4' the hygrometer, with the beard mounted vertically so as to turn the index.

revolving type, recording the revolutions in 10,000's, 1000's and 100's. The direction of the wind was also recorded.* The Society saw the clock on 29 May 1679, but there is no detailed description.† On 2 April 1684 Hooke was told to help his assistant 'to reduce into writing some of the first papers marked by the weather-clock, that thereby the Society might have a specimen of the weather-clock's performances before they proceed to the repairing it'.‡ There thus seems no doubt that the clock worked, if but intermittently. But what a project even if it had not worked!

Hooke can safely claim to be the founder of modern meteorology. Typical of his quantitative spirit is his calculation, in 1662, of the height of the atmosphere from Boyle's law. He worked out that at 25 miles the pressure should be 0.02 atmosphere, which is, of course, much too high, but his conception of pressure decreasing exponentially is modern in spirit. The height of the atmosphere, he said, seemed to be indefinite, but might be many hundreds of miles. He made extensive barometric observations, with the object of finding the connexion between barometric pressure and weather; for instance, in 1677 he said that in sixteen years' observation he had found the quicksilver always to be very low, and to fall to that position rapidly, before any considerable storm, and therefore he hoped that the instrument would be of great use at sea, to foretell storms. He was perfectly clear as to the general polar circulation, saying that the air near the earth moved from the poles to the equinox and the higher parts of the air from the equinox to the poles. But perhaps what is most striking is the way in which he proposed systematic meteorological observations by means of his instruments. The table from his article on 'A Method of Making a History of the Weather', published in 1667 in Sprat's *History of the Royal Society*, is astonishing (figure 5). He is further quite clear that the heat of the sun is conveyed to the air by radiation falling on the earth. He thought of the weather as a matter of physical laws applied to the earth and its atmosphere, the movements of the air as governed by the absorption of solar heat and by the rotation of the earth, in quite a modern spirit. Here we may note that in 1678, pointing out that the rotation of a planet would lead to a spheroidal shape, he remarked on the influence of the diurnal motion of the earth 'which compounded with that of the moon he conceived to be the cause of the tides'.§ Proudman says of Newton's work on the tides: 'The only important factor which he did not mention is the dynamical effect of the earth's rotation.' Of course, there is no comparison between a considered mathematical theory and a shrewd qualitative generalization, but the remark does illustrate Hooke's exceptional acuteness.

Hooke constructed the first modern astronomical instruments. In his keen analysis of the accuracy attainable with astronomical instruments designed for measuring angles, Hooke at once met with the question of the resolving power of the human eye, which he was certainly the first to discuss. He goes into the matter in detail in his Cutlerian lecture *Animadversions on the First Part of the Machina Coelestis*, published in 1674. Hevelius, the author of the *Machina Coelestis*, had built

* *Philosophical Experiments and Observations of the late Eminent Dr Robert Hooke. Publish'd by W. Derham, F.R.S.*, 1726, p. 41.

† Birch, III, 487.

‡ Birch, IV, 277.

§ Birch, III, 390. See also Waller, xx.

an observatory at Danzig, and this he had equipped at great expense with magnificent open-sight instruments, which he used for all his measurements of angular distance. Those instruments were large and of beautiful workmanship: one is represented in figure 6. Hooke's criticism was that open-sights were inherently incapable of the accuracy that could easily be obtained with telescopic sights, at, he said, a tenth the cost. 'I am', he said, 'not satisfied that his Instruments are capable of making Observations more accurately than those of *Ticho*, though 'tis possible that they

A S C H E M E

At one View representing to the Eye the Observations of the Weather for a Month.

Days of the Month and place of the Sun. Remarkable hours.	Age and Sign of the Moon at Noon.	The Quarters of the Wind and its strength.	The Degrees of Heat and Cold.	The Degrees of Dryness and Moisture.	The Degrees of Pressure.	The Faces or visible appearances of the Sky.	The Notable Effects.	General Deductions to be made after the tide is fitted with Observations: As,
14 H 12.46	4 8 12 9. 46.	W. 3 12 31	2 9 16	1 2 1/2	5 29 8	Clear blew, but yellowish in the N. E. Thunder, far Clowded toward the S. Checker'd blew.	A great dew. Thunder, far to the South. A very great Tide.	From the last quart. of the Moon to the change the weather was very temperate but cold for the season; the Wind pretty constant between N. and W.
15 II 13.40	8 4 6 24. 51.	N.W. 3 4 28	9 4 17	2 8 1/2	29 1/2 10	A clear Sky all day, but a little checker'd at 4. P.M. at Sun-set red and hazy.	Not by much as yesterday. Thunder in the North.	A little before he last great Wind, and till the Wind rose at its highest, the Quicksilver continued descending till it came very low; after which it began to ascend,
16 II 14.37	10 at 7. 25' A.M. II 10. 8.	N. Moon. S. 1 10 8.	1 10 8.	1 10 8.	1 10 8.	Overcast and very low-lying.	No dew upon the ground, but very much upon Marble Stones, &c.	&c.

FIGURE 5. From Hooke's paper on 'A Method of Making a History of the Weather'.

may do it with somewhat less trouble and inconvenience', Tycho being, of course, Tycho Brahé. He gives half a minute as the limit of accuracy possible in measuring distances in the heavens with the naked eye 'and hardly one of a hundred can distinguish a minute'. The resolving power of the eye for two near bright objects is taken to-day as just under a minute (50 sec.), and Hooke demonstrated this before the Royal Society on 15 January 1673/4. The ability to set on an object may be slightly better than this—I do not know if it has been measured—but it certainly has a limit of about this magnitude. Hooke points out that it is useless to graduate an instrument with open-sights to take altitudes to seconds—quite apart, as he says, from the fact that reading to seconds does not mean accuracy to seconds—which Hevelius claimed to do by means of a vernier. He then discusses, in a quite modern

manner, the advantages of telescopic sights and describes in detail a mural quadrant with a fixed, horizontal telescope, for sighting on reference marks, and a movable telescope attached to the rotating arm. The telescopes (figure 7) are

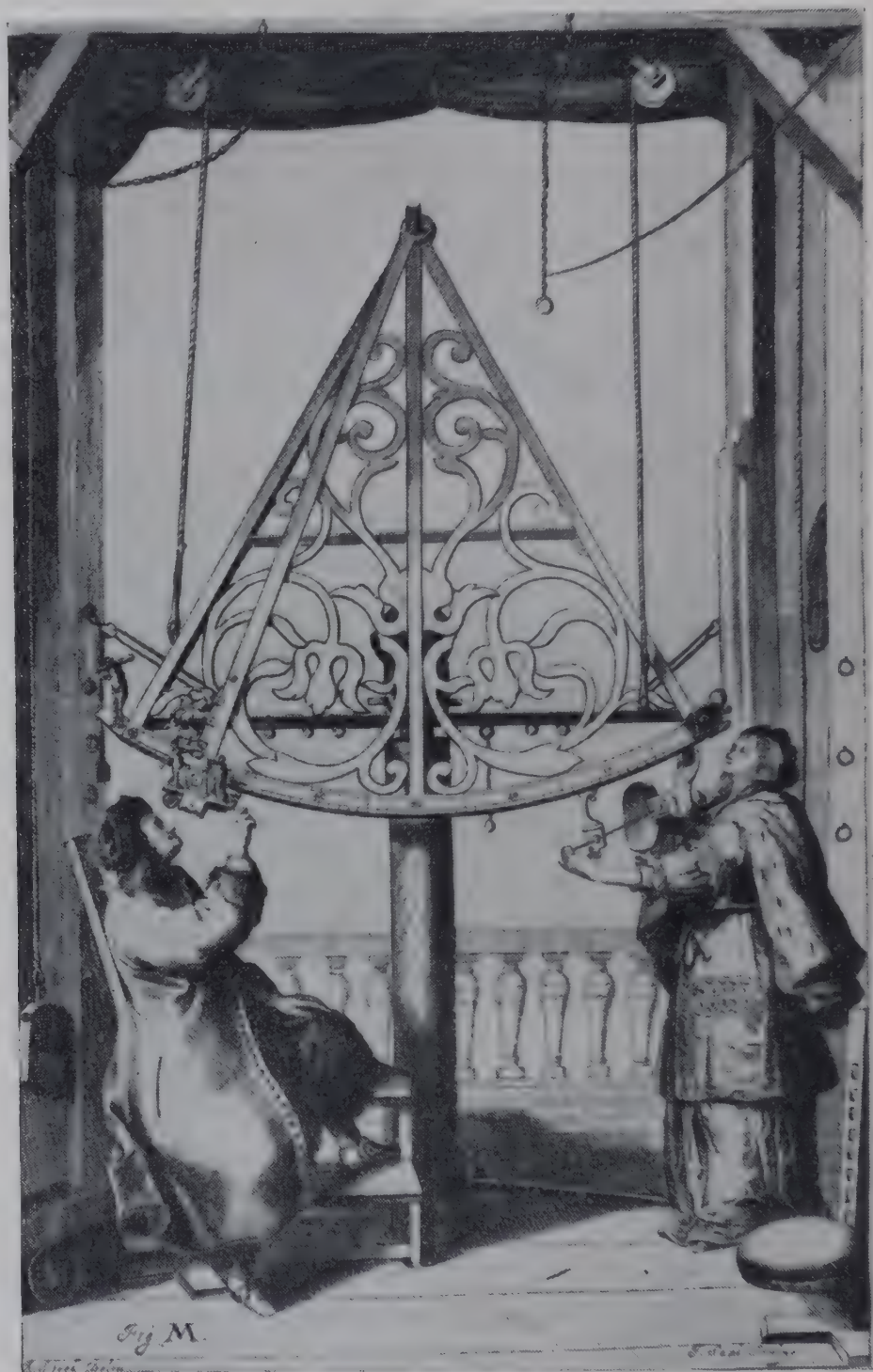


FIGURE 6. One of the open-sight instruments of Hevelius, contemporary with Hooke's telescopic sights.

furnished with cross-wires, made of spider's web or single silk fibre, at the focus of objective and eyepiece; furthermore, by metal reflectors, things are so arranged that an observer looking horizontally can see both objects on the cross-wires and bring them into coincidence by turning the moving telescope. Apart from telescopic sights, this is an anticipation of the Hadley (or Newtonian) sextant. As regards the

quadrant, Hooke describes how this may be accurately graduated by means of a screw—the first dividing engine (see figure 8). Incidentally, he describes the new invention of the universal joint or Hooke's joint (figure 9). He also describes a spirit

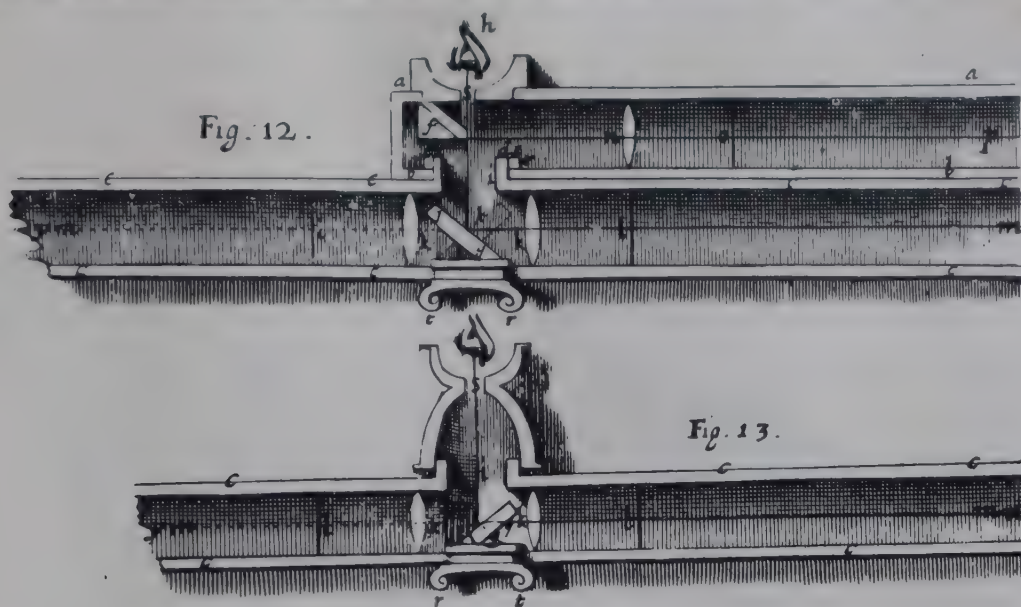


FIGURE 7. Hooke's telescopic sights, fixed and rotating, for his mural quadrant.

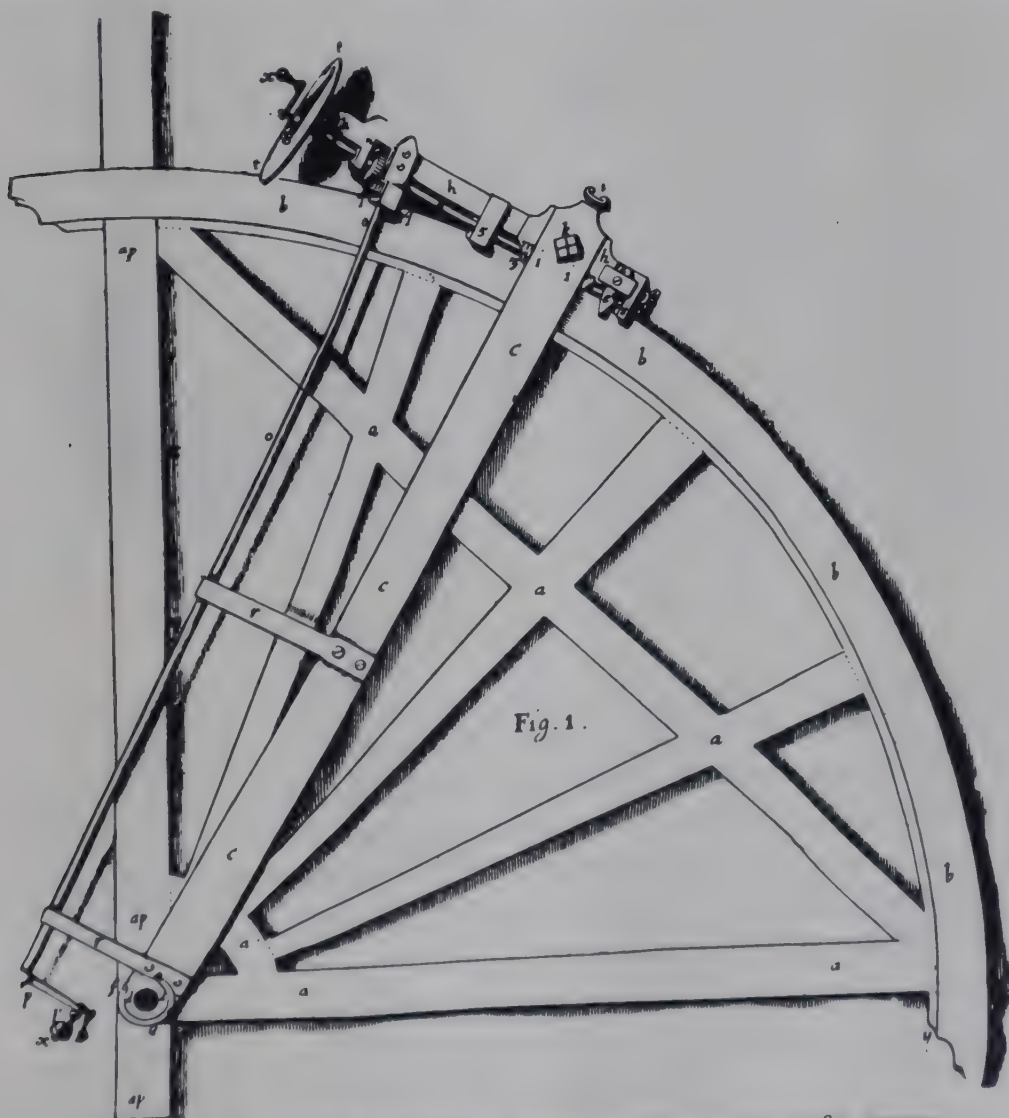


FIGURE 8. Hooke's quadrant, graduated by means of a screw.

level, but here he had been anticipated by Thevenot. Anyone who contrasts Hooke's instrument with those of Hevelius will see a difference of kind; one is completely modern in spirit, with a careful discussion of the limits of accuracy and of experimental error, the other is a much improved version of the instruments of the ancients.

But this is by no means all in this one section of the Cutlerian lectures. Hooke describes the first clock-driven telescope (figure 10), although there seems to be no evidence that he ever made such an instrument. It is driven by a conical pendulum

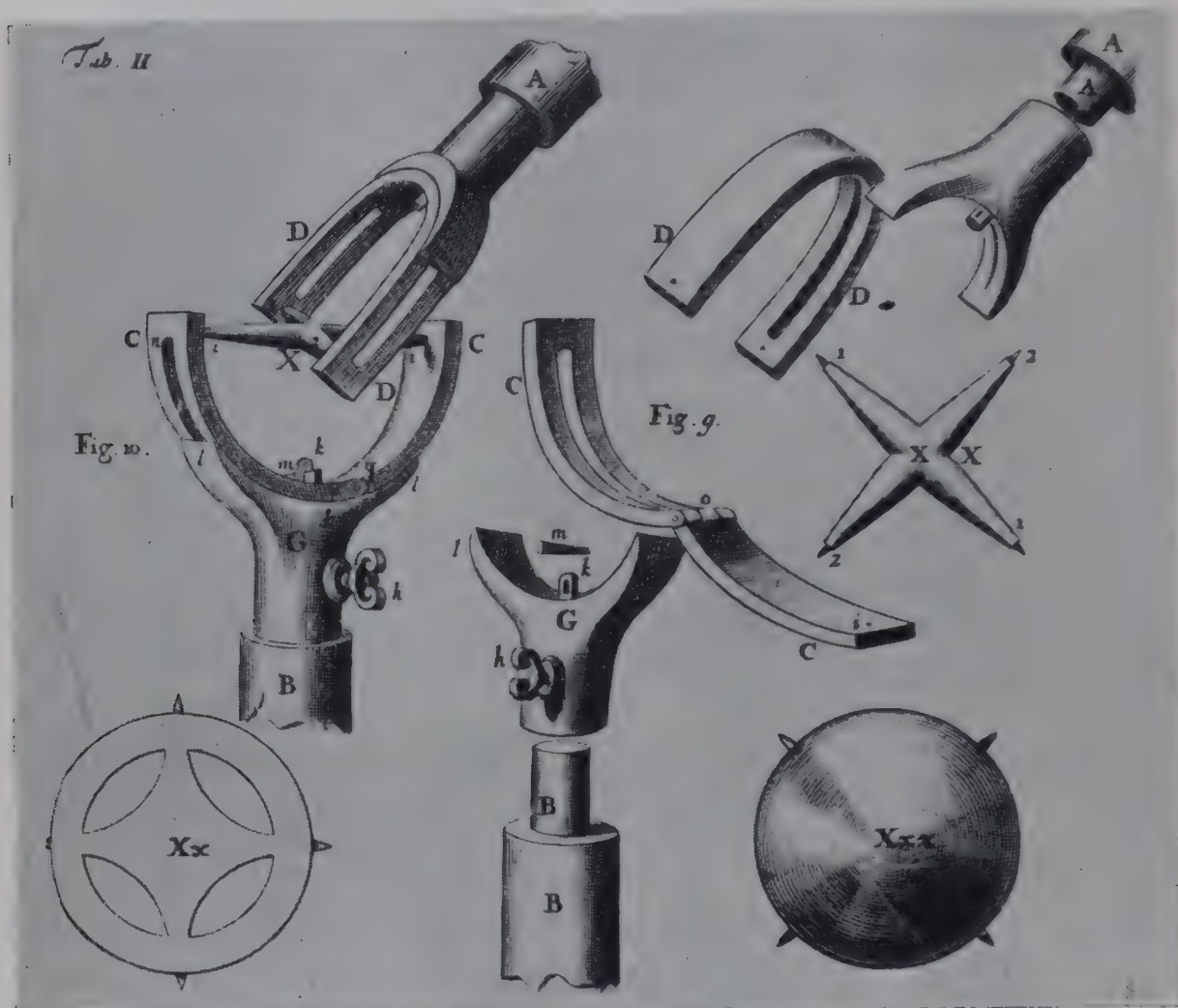


FIGURE 9. The universal joint or Hooke's joint.

(or circular pendulum, as Hooke called it), and the details given certainly read as if set down in the light of experience in the construction. As regards the conical pendulum, Hooke complains that Huygens, in his *Horologium Oscillatorium* (1673), gives a description of it without acknowledging that he, Hooke, invented it. This claim of his appears to be well-founded, for in Birch's *History* we find that on 23 May 1666 Hooke did discuss the conical pendulum. Casually, in the *Animadversions*, he gives the first invention of a spiral gear.

With regard to horology, I have already referred to the spring-controlled balance wheel, which I am satisfied to attribute to him, but about which there has been controversy. The invention of the anchor escapement is generally ascribed to Hooke, but here I can find no really satisfactory evidence for the attribution.

He constructed the first reflecting telescope of the Gregorian pattern, using spherical mirrors, after Newton had made one of the pattern to which he gave his name. It appears from Dr Pell's collection in the British Museum,* as well as from what Hooke himself says in his *Animadversions on Hevelius*, that Hooke invented about 1670 an engine for multiplying and dividing.† Samuel Morland published the account of his calculating machine in 1672. As we have seen, Hooke was responsible for the construction of Boyle's first pump.

Let us now look at a few of his minor inventions. He can certainly claim the iris diaphragm; the description in Birch, under the meeting of 27 July 1681, is precise: 'Mr Hooke showed his new-contrived aperture for long telescopes, which would

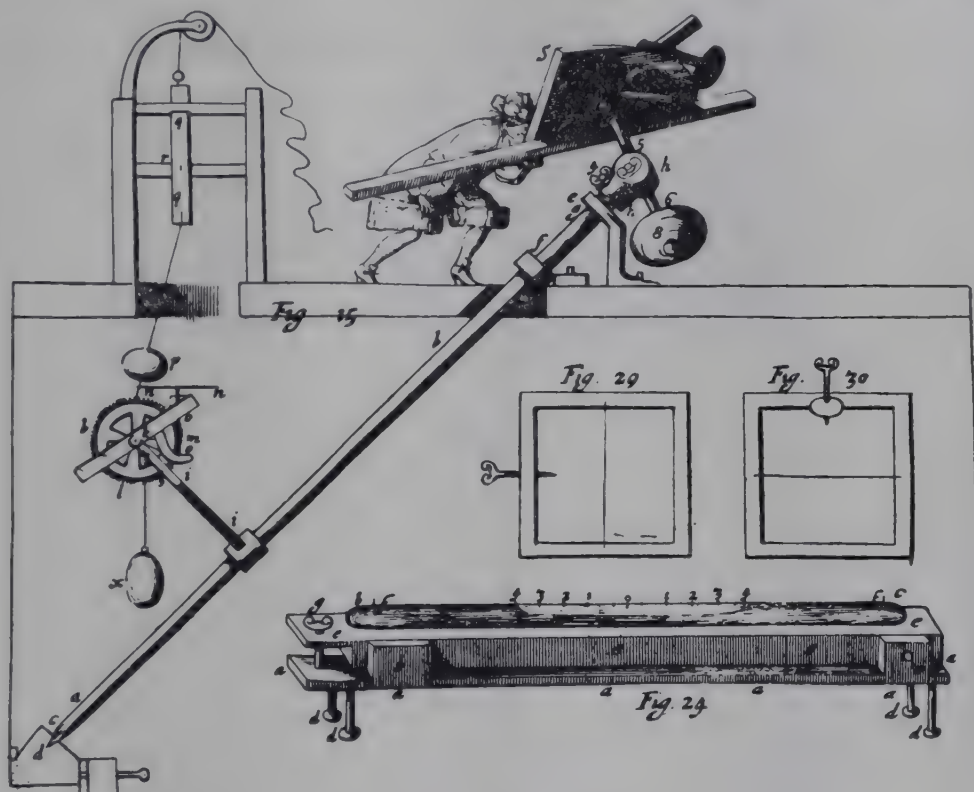


FIGURE 10. Hooke's figure of a clock-driven telescope.

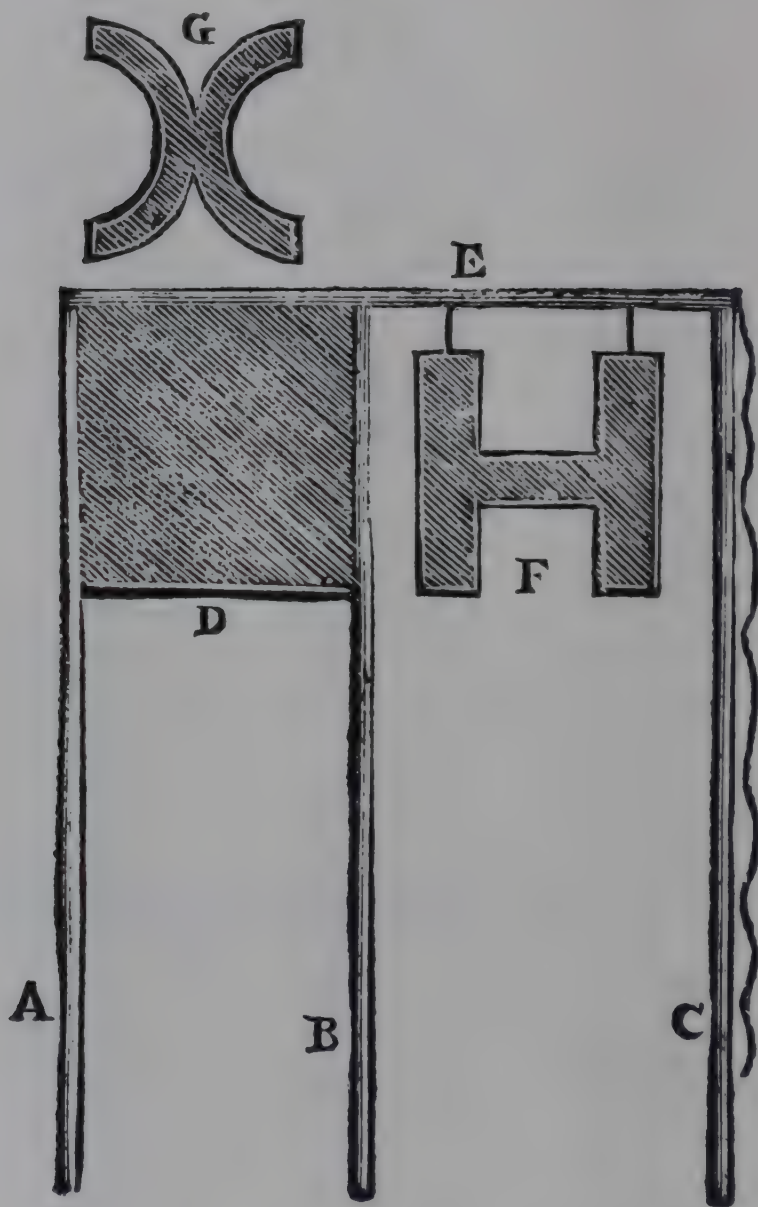
open and close just like the pupil of a man's eye, leaving a round hole in the middle of the glass of any size desired; which was well approved of.' Apparently Hooke used to have the notion of some such device, make it in a morning, show it at the Royal Society, and forget it. At the same meeting he showed his improved helioscope, 'which would exhibit the figure of the sun very perfectly'. He invented ingenious machines for sounding the sea, and one for bringing up samples of sea water from the depths. He first used immersion lenses for the microscope, pointing out their advantage. He made the first refractometer for liquids.

His glimpses into the future are astonishing. I have already referred to his conclusions about flying. He says: 'The way of flying in the Air seems principally impracticable, by reason of the want of strength in humane muscles; if therefore

* See *A Brief Account of the Life, Writings and Inventions of Sir Samuel Morland*, by James Orchard Halliwell, 1838, p. 13.

† See also Waller, xix.

that could be supplied, it were, I think, easie to make twenty contrivances to perform the office of Wings.' He therefore attempted to make artificial muscles, using a chain of small bladders, by blowing into which an immediate contraction could be produced. This does not appear to have been taken further. In 'a conjecture,



The Sentences, to be exprefs'd by one Character, may be such as these, in Fig. 2.

O I am ready to communicate.)(I am

ready to observe. (I shall be ready presently.

) I see plainly what you shew.) Shew the

last again. (Not too fast. Shew faster. Answer me presently. Dixi. Make Hastie to communicate this to the next Correspondent. I stay for an Answer; and the like.

FIGURE 11. Hooke's figure of his optical telegraph; under it is shown a passage in which he describes his characters designed to represent sentences frequently needed.

that it may perhaps be possible to spin a kind of artificial Silk out of some glutinous substance that may equalize natural Silk', he says, 'And I have often thought, that probably there might be a way found out, to make an artificial glutinous composition, much resembling, if not full as good, nay better, then that Excrement, or whatever other substance it be out of which, the Silk-worm wire-draws his clew. If such a composition were found, it were certainly an easie matter to find very quick ways of drawing it out into small wires for use. I need not mention the use of such an Invention, nor the benefit that is likely to accrue to the finder, they being sufficiently obvious. This hint therefore, may, I hope, give some Ingenious inquisitive Person an occasion of making some trials, which if successfull, I have my aim, and I suppose he will have no occasion to be displeas'd.' It is possible that he anticipated the stethoscope. He says: 'There may also be a possibility of discovering the internal motions and actions of bodies...whether animal, vegetable, or mineral, by the sound they make; that one may discover the works performed in the several offices and shops of a man's body, and thereby discover what instrument or engine is out of order...I have been able to hear very plainly the beating of a man's heart.' He certainly invented the optical telegraph, subsequently developed, at the end of the eighteenth century, by Claude Chappe. He pointed out that the telescope enabled the signals to be read from a great distance, discussed conditions for good visibility and devised a system of signals for day and night use, all in a very practical spirit, as the passage reproduced in figure 11 will attest. These are a few examples of his sagacious anticipation.

If we turn from particular instruments and inventions to Hooke's general contributions to science, Hooke's law occurs at once to everybody. It was published in 1678 in *Lectures de Potentia Restitutiva or of Spring*, but had been announced two years earlier in the form of an anagram, *ceiinoosssttuu*; letters which, rearranged, make the words *ut tensio sic vis*—'as the extension, so the force', or, as we say nowadays, strain is proportional to stress. The law is illustrated in a number of ways—by the wire under axial tension, the helical spring, the spiral spring of the watch balance, the wooden cantilever (see figure 12). Hooke also quotes what is generally known as Boyle's law as a further example. Accurate measurements of the relation between volume and pressure of air he had published thirteen years earlier in the *Micrographia*; there he says that he first carried out experiments of this kind in 1660 and repeated them in 1661, when, having heard Townley's hypothesis, he modified the method of recording, and exactly verified that 'the Elater of the Air is reciprocal to its extension, or at least very neer', the elater being, in modern parlance, the elasticity. Thus it was Hooke who really discovered Boyle's law. That this is so is, I think, substantiated by two arguments: first, that Hooke, who always expressed the greatest veneration for Boyle, would never have published this in 1665 if it was likely to give pain to, or be disputed by, Boyle; and secondly, that Boyle says in 1662* 'wherein I had the assistance of the same person that I took notice of in the former Chapter...whom I had rather make mention of on this occasion, because when he first heard me speak of Mr Townley's suppositions about the proportion wherein Air loses its Spring by Dilatation, he told me he had the

* *A Defence of the Doctrine touching the Spring and Weight of the Air*, p. 64.

year before...made observations to the same purpose, which he acknowledged to agree well with Mr Townley's theory.' Thus Boyle's account and Hooke's are substantially the same. L. T. More, in his life of Boyle, says: 'The specific experiments point to Boyle, but the method of attack, which involved the physical and mechanical properties of the air, shows the influence of Hooke', and points out that after Hooke left him Boyle never did any quantitative work.

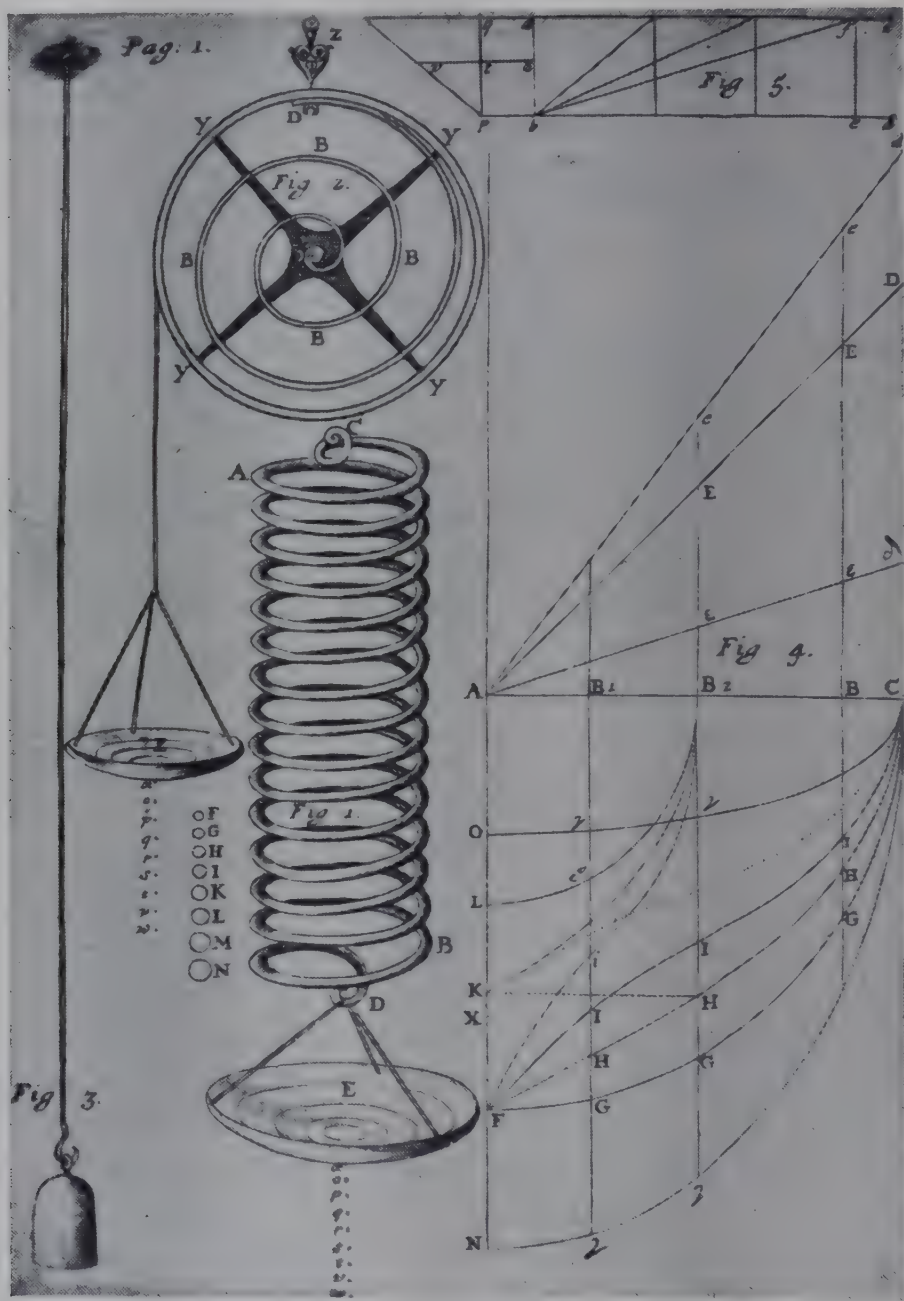


FIGURE 12. The plate which illustrates Hooke's *De Potentia Restitutiva* or of Spring.

The *Potentia Restitutiva* contains other things of major importance besides Hooke's law. Hooke shows that any vibration in which the restoring force is proportional to the displacement must be isochronous, a capital discovery. He points out that the restoring force of a bent beam is due to the extension of the under part and the compression of the upper part. Further, here and elsewhere, he develops a crude kinetic theory of matter. He supposes that all the particles of the universe are in

continual motion: 'Two or more of these particles joyned immediately together, and coalescing into one become of another nature, and receptive of another degree of motion and Vibration, and make a compounded particle differing in nature from each of the other particles'*—not a bad definition of molecules. He further supposes a universal ether which is the medium that 'conveys all Homogenous or Harmonical motions from body to body'. In a solid the particles touch at their extreme librations; in a liquid they do not, and the subtle ether gets between them. He says that all the particles of a solid are kept together by the motions of the surrounding ether and tend to fly apart by virtue of their own motions, and that 'according to the prevalence of the one or the other is the body more or less fluid or solid'. Air 'consists of the same particles single and separated, of which water and other fluids do, conjoined and compounded', only they are very widely spaced—'its Vibrative Spaces exceeding large, comparative to the Vibrative Spaces of other terrestrial bodies'.† He mentions a million a second as the kind of frequency he has in mind. On the basis of this general kinetic theory he explains the elasticity of all bodies. Jeans‡ says of this work: 'He suggested that the elasticity of gases resulted from the impact of hard independent particles on the substance which enclosed it, and even tried to explain Boyle's law on this basis.' When we turn to consider Hooke's thoughts on heat we shall again find this conception of universal vibration.

Hooke was the first to speculate in a modern way about crystal structure. In the *Micrographia* he points out that the forms in which alum and rock salt crystallize can be built up of spherical particles, and he gives examples, in two-dimensional drawings (figure 13), saying that he has built up models with 'globular bullets' of all the regular figures commonly met with in crystallization. He cites alum and rock salt—cubic crystals—in particular, but outlines a general programme concerning crystallization which he would like to pursue if he had time and assistance, ending: 'So that knowing what is the form of Inanimate or Mineral bodies, we shall be the better able to proceed in our next Enquiry after the forms of Vegetative bodies; and last of all, of Animate ones, that seeming to be the highest step of natural knowledge that the mind of man is capable of.' Our biophysical friends will agree.

Hooke was certainly one of the first, if not the first, to draw attention to thermal expansion as a general property of matter. 'This property of Expansion with Heat, and contraction with Cold, is not peculiar to Liquors only, but to all kinds of solid Bodies also, especially Metals.' As regards the nature of heat, it was, he said in the *Micrographia*, 'nothing else but a very brisk and vehement agitation of the parts of the body' and again 'Heat is a property of a body arising from the motion or agitation of its parts'. He points out that if you rub any hard bodies together they grow hot, and says that a violent enough motion will melt a body, quoting, as an example, sparks struck from steel, which he showed microscopically to be particles molten

* *Potentia Restitutiva*, p. 9.

† *Potentia Restitutiva*, pp. 15, 16.

‡ Jeans, *An introduction to the kinetic theory of gases*, 1940, p. 3. See further Tait, 'Hooke's anticipation of the kinetic theory', *Proc. Roy. Soc. Edinb.* 16 March 1885. *Tait's Collected Works*, II, 122.

into a spherical form. Bodies that melt easily have parts unapt to cohere, so that a small degree of agitation keeps them fluid. Much else that he writes has a very modern sound, much more so than the speculations of the eighteenth century. The same we find when we turn to his work on combustion.

In the theory of combustion and respiration Hooke rendered services which have been often overlooked, although Thomas Thomson, in his *History of the Royal Society*, 1812, says: 'Mayow, in his Essays, published at Oxford about ten years after the *Micrographia*, embraced the hypothesis of Dr Hooke, without acknowledgement; but clogged it with so many absurd additions of his own, as greatly to obscure its lustre and diminish its beauty.' Ripe scholars like Dr McKie have of

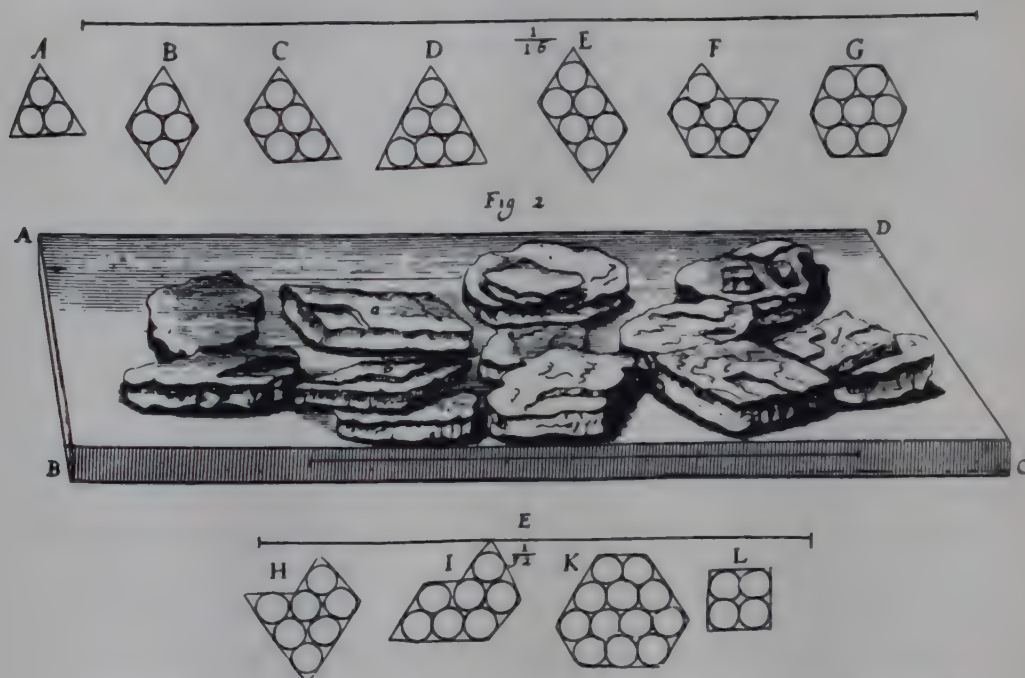


FIGURE 13. Hooke's representation, by means of spherical bullets, of simple crystalline forms shown by alum.

recent years given balanced and appreciative accounts of Hooke's work in this field, but in the average history of chemistry Hooke's contributions are neglected. The record of the subject is confused and complicated, and it is possible that to abbreviate it is to introduce an artificial simplicity, but it must be attempted.

It was Boyle who first showed, with the pump which Hooke made for him, that reducing the pressure reduced the time that a candle could burn or a mouse or bird live in a given vessel. Hooke was responsible for the experiment, described in histories of medicine, in which a dog, of which the thorax and belly were displayed, was kept alive for a long time by air supplied through the windpipe by a pair of bellows, showing that the air, and not the muscular motion of the ribs, was the essential. In the *Micrographia* he described a crucial experiment in which he showed that wood heated in a closed iron vessel became charred, but would not burn in the absence of air. His theory of combustion was that combustible—sulphureous, as he called them—bodies dissolved in the air by virtue of a substance present in the air that was 'like, if not the very same, with that which is fixed in *Salt-peter*, which by multitudes of Experiments that may be made with *Salt-peter*, will, I think, most evidently be demon-

strated'. This substance was used up in combustion, so that a fresh supply of air, as, for instance, that produced by bellows, was necessary for continuing combustion. Thus he clearly realized the functions of oxygen without naming it further than by saying it was copiously contained in saltpetre, or, as he himself said very plainly, concerning respiration, air contains 'a kind of nitreous quality which being spent made the air unfit for respiration'. He further said that the function of respiration was to introduce from the air into the blood something essential to life and to remove from it something noisome that was discharged back into the air. Experiments to find out if the volume of air changed due to combustion were not decisive; it was here that Mayow, performing the experiment over water, made an advance. But I hope that I have indicated that, before his time, Hooke had arrived at the essentials of the true theory of combustion and respiration. Incidentally in 1671

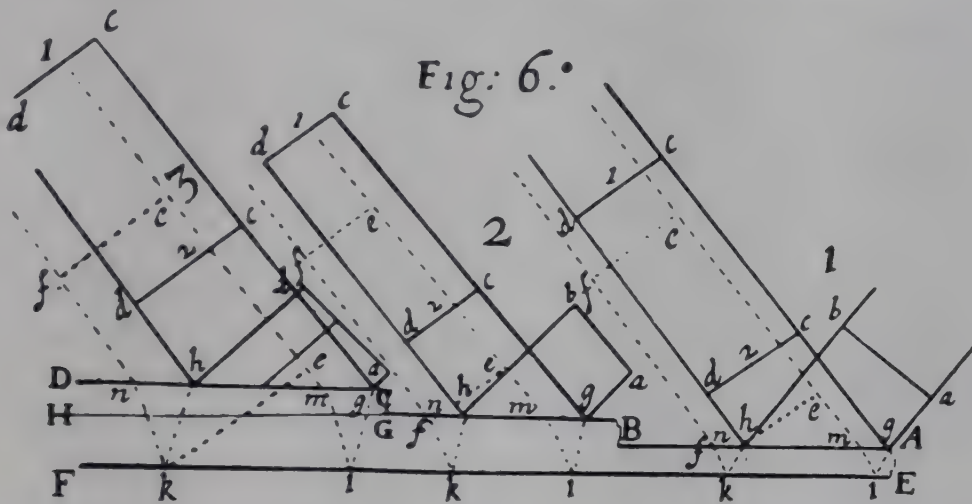


FIGURE 14. Hooke's diagram to represent the way in which the colours of thin plates are formed.

Hooke tried some experiments, with himself as the subject, on breathing at reduced pressure. The experimental compartment consisted of two tuns, one within the other, the space between being filled with water. He managed to retain the pressure at a quarter of an atmosphere below normal and stayed in this rarefied air for a quarter of an hour; now and then fresh air was let in. He felt little inconvenience except about the ears, as anyone who has flown at about 8000 feet can well believe. These must be the first experiments ever carried out on the effect of artificial changes of air pressure on human comfort.

As regards light, the *Micrographia* contains, first, a discussion of the colour of thin plates which shows Hooke's usual acuteness of observation and perception of essentials. Having found that the colour of flakes of mica only appeared when the thickness was very small, he produced colours with thin films of air between glass plates and observed the changes of colour that took place as the plates were pressed closer together. He was quite clear that the colours of thin plates were produced by the light reflected from the back of the plate acting together with that reflected from the front—his diagram (figure 14) would do for to-day. It is only his notion of the mechanism of interaction that is obscure. He treated the colours of thin bubbles

of various kinds. He observed with a lens pressed on glass the rings which bear Newton's name. He said that the colours of steel were produced by a thin lamina of 'vitrified' (oxidized) steel, and conjectured that the hardness of tempered steel was due to 'vitrified' substances interspersed through the pores of the steel. What I wish to stress here is not the possibility of interpreting this as a foreshadowing of modern theory but that Hooke is attempting what, I believe, nobody had considered before, the explanation of the change of properties of metals by heat treatment on strictly physical and structural lines. 'Those metals which are not so apt to vitrify, do not acquire any hardness by quenching in water, as Silver, Gold, &c.' I should like to draw your attention to the fact that round about the beginning of 1676 Newton often refers to the *Micrographia*, and in his discourse read before the Society on 10 February 1675/6 compliments Hooke on his observations on thin plates and says 'which I have not scrupled to make use of so far as they were for my purpose'. And on 5 February 1675/6 he wrote to Hooke: 'You have added much several ways, and especially in taking the colours of thin plates into philosophical consideration. If I have seen further it is by standing on the shoulders of Giants. But I make no question but you have divers very considerable experiments besides those you have published, and some it's very probable the same with some of those in my late papers. Two at least there are which I know you have observed, the dilatation of the coloured rings by the obliquation of the eye, and the apparition of a black spot at the contact of two convex glasses and at the top of a water bubble; and it's probable there may be more, besides others which I have not made.' This makes his ungenerous lack of reference to Hooke in the *Opticks* all the more deplorable.

Hooke's theory of light is that it is produced by a very rapid vibratory motion, of very small amplitude, in the luminous body, which need not, however, be hot, since rotten flesh and wood give out light, but are cold. The light is carried by an all-pervading transparent homogeneous medium—the ether, as all the classical writers on light called it. In the free ether the 'orbicular pulse' is at right angles to the ray—whether this is to be called a transverse wave is a matter which can be lengthily discussed. Like others, Hooke supposed the disturbance to travel faster in the refracting medium, which, by a kind of Huygens construction—long before Huygens's exposition—became a beam with the 'pulse' oblique to it in the medium. It was a happy guess of Hooke's that the vibrations were transverse, but this obliquity in the medium spoilt it. His theory of colour, to which he attached considerable importance, is obscure; he opined that it was caused by refraction, but says that red is the impression produced by an oblique and confused pulse whose strongest part precedes and whose weakest follows, while blue is the impression produced by a like oblique and confused pulse whose weakest part precedes and whose strongest follows. What he meant by 'confused' and by the weakest and strongest parts I do not know. For him red and blue were the primaries and the other colours were produced from them. This theory of colour is not the happiest part of Hooke's theories, and his ill-tempered attacks on Newton's masterly work with the prism is one of the very few cases where he did not appreciate a great discovery.

A passing reference must be made to Hooke's work on diffraction, nine years subsequent indeed to that of Grimaldi, but almost certainly made in ignorance of Grimaldi's work. Hooke speaks of a new property of light not observed, that he knew of, by any optical writer and Priestley* agrees that Hooke's experiments are so different that it is extremely probable that they were completely original. Hooke describes both the deflexion of the light and the colours produced by it. He compared it to the straying of sound into the quiescent medium; Newton, however, said that it was only a new kind of refraction, caused by a variation in the density of the ether near the edge of the solid body. Newton likewise pointed out, no doubt with some satisfaction, that he had found the phenomenon in Faber's *De Lumine* and that Faber had taken it from Grimaldi. Hooke, however, with his usual acuteness, hit upon the allied phenomenon of diffraction in sound for comparison.

As regards sound, Hooke did not do very much, but what he did was, as usual, an anticipation of much that was done again later. 'He showed an experiment for making musical and other sounds by the help of teeth of brass wheels, which teeth were made of equal bigness for musical sounds, but unequal for vocal sounds',† which is Savart's wheel (date 1820) with an added refinement. He carried out experiments on the vibrations of an inverted glass bell full of water; he stated that the motion of the glass was from oval to circular 'and presently changed into an oval the other way', which he discovered by the rising of water in the glass. 'Sir Christopher Wren coming in said, that the glass would vibrate much stronger, being struck on the edge with a viol-bow.'‡ They then obtained the vibration in six and in eight segments; the four and the eight were octaves and the six and the four were fifths. These observations on bells are usually attributed to Chladni, who carried out his experiments in 1787, using water, just as did Hooke. Hooke also did something in the way of obtaining Chladni figures with flour as the powder, 'it being', says the account, 'manifest by this experiment, that as every different stroke makes a different sound, so the making of a different impression upon the flour gives it as many several motions';§ or, in other words, the flour indicated a different mode of vibration for each note. He also found the frequency of a given musical note with a monochord,|| a matter which he duly discussed with Samuel Pepys. 'Discoursed with Mr Hooke about the nature of sounds, and he did make me understand the nature of musicall sounds made by strings, mighty prettily; and told me that having come to a certain number of vibrations proper to make any tone, he is able to tell how many strokes a fly makes with her wings, those flies that hum in their flying, by the note that it answers to in musique, during their flying. That, I suppose, is a little too much refined; but his discourse in general of sound was mighty fine.'

* *History of Vision, Light and Colours*, 1772, p. 172. The reference which Priestley there gives to Birch is wrong; it should be Birch, III, 194, which in its turn refers to the *Posthumous Works*, p. 186.

† Birch, IV, 96. See also Waller, xxiii.

‡ Birch, IV, 46.

§ Birch, II, 475.

|| Birch, I, 446, 451.

Hooke tells us in the *Micrographia* that he had experimented on the propagation of sound through wires over a very considerable distance and adds that it travelled in the wire incomparably swifter than in the air—it actually does travel some fifteen times as fast in an iron wire, say, as in air. He seems to have considered this as a possible method of communication.

Hooke is, of course, one of the great classical microscopists, whose name would live with those of Malpighi; Grew, whom he probably inspired; Swammerdam; and Leeuwenhoek had he left behind only his microscopic observations. His pictures of moulds, of moss, of the nettle, and of insects and their parts, such as the bee's sting and the fly's eye, have been objects of admiration ever since they were executed. The figures of the flea and the louse were long famous. I have a little book called *The Wonders of the Microscope*, published in 1811, which figures a Culpepper microscope of the type made by Harris in about 1780, but is illustrated entirely by Hooke's figures, apparently printed from the original plates. His representation and discussion of the structure of a feather was long the standard one.* His investigation of the microscopic structure of cork led him to discuss the cellular structure of plants, which he was the first to recognize; it is to him that we owe the introduction of the word 'cell' in its biological sense. Perhaps here mention may be made of his work on the life cycle of the gnat, which is a biological investigation of great significance.

As a geologist and early evolutionist Hooke was far in advance of his time. For his contemporaries geology was mainly a subject for fanciful speculations as to the origin and structure of the earth, which made little use of observation and material evidence, but much use of biblical legend and of the teaching of the church in general. For instance, two contemporaries of Hooke were Thomas Burnet and William Whiston: Burnet published in 1681 *Telluris Theoria Sacra*,† which appeared later in English as the *Theory of the Earth*; here he maintained that the earth was like a gigantic egg, the shell of which at the time of the Deluge—the great geological upheaval—was split, with the consequence that internal waters gushed out, and so on: it was the wickedness of mankind that led to the disaster. William Whiston, in his *New Theory of the Earth*, which appeared in 1696, supposed that the rotation of the earth began with the fall of man, and that, as with Burnet, at the time of Noah the central abyss of waters broke through the crust, a comet having something to do with it. Hooke's geology, which is to be found in the *Posthumous Works*, stands out in sharp contrast to these fanciful romances. For him fossils were a definite record of past life, and not mere curious freaks of nature. He made excellent drawings, of which those of the ammonites (figure 15) are typical—W. N. Edwards says that they are not always equalled to-day—and argued forcibly that fossils are either organisms themselves turned to stone or else impressions left by such organisms. The fact that remains of marine creatures were found high above the sea he put down to the upheavals of the earth's surface which have taken place: it should be noted that he used the term 'earthquakes' to denote any

* To reproduce here any of the figures mentioned, on the reduced scale that would be necessary, would not do them justice. Reference should be made to the originals.

† There was a second volume of the Latin work in 1689, and of the English in 1690.

serious displacements of the earth's surface. He discounts Noah's flood, so popular with his contemporaries as an explanation of the past life of the globe, pointing out that it cannot have lasted long enough for the production of the shells which are found high on dry land—'besides the quantity and thickness of the Beds of Sand with which they are many times found mixed, do argue that there must needs be a much longer time of the Seas Residence above the same, than so short a space



FIGURE 15. Hooke's drawings of ammonites.

can afford'. His general opinion, so much in advance of his time, let him give in his own words, words that are a good example of his beautiful, clear and direct prose: 'I do therefore humbly conceive (tho' some possibly may think there is too much notice taken of such a trivial thing as a rotten Shell, yet) that Men do generally too much slight and pass over without regard these Records of Antiquity which Nature have left as Monuments and Hieroglyphick Characters of preceding Transactions in the like duration or Transactions of the Body of the Earth, which are infinitely more evident and certain tokens than any thing of Antiquity that can be fetched out of Coins or Medals, or any other way yet known, since the best of those ways may be counterfeited or made by Art and Design, as may also Books, Manuscripts and Inscriptions, as all the Learned are now sufficiently satisfied, has often been actually practised; but those Characters [fossil shells] are not to be counterfeited by all the Craft in the World, nor can they be doubted to be, what they appear, by any one that will impartially examine the true appearances of them: And tho' it must be granted, that it is very difficult to read them, and to raise a *Chronology* out of them, and to state the intervalls of the Times wherein such, or such Catastrophies and Mutations have happened; yet 'tis not impossible, but that, by the help of those joined to other means and assistances of Information, much may be done even in that part of Information also.'* I myself well remember as a little boy finding fossil shells on the cliffs at Alum Bay in the Isle of Wight, high above the Needles, a couple of miles from Hooke's birthplace, and I like to think that it was, perhaps, boyhood memories of some such finds that interested Hooke particularly in the matter.

There is no time to refer briefly to more than a small part of his astonishing speculations on cosmological matters. The sun, he says, is solid, encompassed with a bright atmosphere, clouds in which are the sun spots and movements of which give the faculae; all this he argues from observed facts. The light of the sun is due to heat produced by some kind of combustion; sulphur and nitre make a very bright flame, but probably the materials of the sun are much better adapted for producing heat—a conjecture which recent discovery has certainly confirmed! The stars are probably bodies much like the sun. The nearest of them, he deduces from measurement, cannot be nearer than 68,760,000 earth diameters, that is, 35 light days—not far enough, but pretty good for the time. And, he says, other stars are much farther and more powerful telescopes will show still farther and farther stars 'certainly this Material Expansum, part of which we are, must be so great that 'twill infinitely exceed our shallow Conception to imagine'. Again, 'To me indeed the Universe seems to be vastly bigger than 'tis hitherto asserted by any writer.'† Of comets he gives much better pictures than does Hevelius, who spent his whole life observing. He explains the behaviour of the comet's tail by repulsion from the sun, which we know now to exist on account of the pressure of light, while clearly

* *Posthumous Works*, p. 411.

† Sir Harold Spencer Jones, the Astronomer Royal, has kindly pointed out to me that already in 1576 Thomas Digges had stated that greater distance was the reason of the relative lack of luminosity of the fainter stars, saying 'the greatest part rest by reason of their wonderfull distance inuisible vnto vs'. See F. J. Johnson, *Astronomical Thought in Renaissance England*, 1937.

recognizing that the body of the comet is attracted by the sun. 'So that though I suppose the attractive power of the Sun . . . may draw the body towards it, and so bend the motion of the Comet from the straight line, in which it tends, into a kind of curve, whose concave part is towards the sun . . . yet I conceive that all those parts of the Comet which are thus wrought upon by the other, and changed into another state, and are very much rarified, and produce light, are of a clear contrary nature, and recede from the centre of the Sun.' He was the first to observe and record a spot on Jupiter; he watched it for two hours and observed that the planet rotated on its axis, but neglected to determine the period. He pointed out that the rings of Saturn are sensibly more luminous than the planet. He observed spots on Mars and from his observations concluded that the planet rotated in a period of 12 or 24 hours. The true period is 24 hours. Actually Hooke's drawing of the markings on Mars was used in recent times by Proctor to determine this period with great accuracy, as 24 hr. 27 min. 32.2 sec.! In 1665 he discovered that γ Arietis was actually two stars 8 sec. apart—one of the earliest records of a double star. One of his most careful and systematic observations was the attempt to observe stellar parallax with a fixed vertical telescope, recorded in *An Attempt to Prove the Motion of the Earth*. He selected the star γ Draconis as his object and concluded, from its apparent position at different times of year, that there was a detectable parallax.* It was found, however, that the displacement of the star presented strange features that could not be explained by parallax, and it was as a consequence of very careful observations, which confirmed the accuracy of Hooke's, that Bradley in 1727, nearly sixty years later, discovered aberration, namely, that the apparent displacement was a consequence of the finite velocity of light.

As regards theoretical astronomy, Hooke most clearly knew the principles on which the planetary system was to be explained; he was, I think, the first to state plainly that the problem was one of ordinary dynamics, with no mystical element. In 1674 he said that his system of the world depended upon three suppositions: first, that all celestial bodies whatever had a gravitating power towards their centres, the one for the other; secondly, that all bodies continue in a straight line motion except they are bent aside by some force—'describing a Circle, Ellipsis, or some other more compounded Curve Line'. The third supposition was that the attractive power fell off with the distance, but he did not as yet know the law. Of this law he says: 'it is a notion which if fully prosecuted as it ought to be, will mightily assist the Astronomer to reduce all the Celestial Motions to a certain rule, which I doubt will never be done true without it'. On 6 January 1680, however, Hooke wrote to Newton 'my supposition is that the attraction always is in duplicate portion to the distance from the centre reciprocally', which is the inverse square law in the language of the day. On 17 January he wrote: 'It now remains to know the propriety of a curve line (not circular nor concentric) made by a central attractive power which makes the velocities of descent from the tangent line or equal straight motion at all distances in a duplicate proportion to the distances reciprocally taken.' If Hooke could have solved this he would have had the glory of

* '...et c'est bien injustement qu'on l'a soupçonné d'avoir ajusté ses observations sur un système adopté d'avance'. Lalande, *Astronomie* (Troisième Édition), Vol. III, p. 81.

a full explanation of the solar system in terms of universal gravitation. He could not solve it, Newton did. But to have enunciated the problem so clearly was a great advance. Surely he did deserve a mention! Newton had it all already, but how was Hooke to realize that?

We now come to consider Hooke's character. The hardest things have been said of him. He has been depicted as melancholy, mistrustful, miserly, penurious, solitary, jealous of all his contemporaries, secretive, claiming all discoveries for himself. Let me admit at once that he was irritable, that he was a man of strong dislikes, and that he particularly disliked being robbed of credit or cheated of money. He was intolerant of fools. He often complained; he often had something to complain of. His mind was so fertile that he was probably often justified in saying that he had made inventions before others. If he sometimes claimed too much credit, he was often denied any credit, as in the passages with Newton. Here he had the, for him, unspeakable chagrin of having before all his fellows attained clear notions on fundamental problems, only to see the greatest of all men of science carry the matter to the final victorious conclusion which was beyond his power. Especially was this the case in the matter of the planetary motions, but his experiments on light also did much to show the way. As I have pointed out, Newton at one time acknowledged that he had learnt much from Hooke's work, but in the end he gave no thanks. The greater man could, as appears from Hooke's letters, so easily have satisfied him.

Yes, he was bitter and bad tempered on many an occasion. He complains that Oldenburg did not enter up in the Minute Book his discourses before the Royal Society: in his Diary he calls him 'a lying dog', 'treacherous and a villain'. Well, there is clear proof that Oldenburg did neglect his work. Pepys enters in his Diary that on 21 February 1665/6 he went with Lord Brouncker to Gresham College and heard a good lecture of Mr Hooke's about the trade of felt-making, 'very pretty'. We have Hooke's notes of the lecture, but there is no record of it in the minutes. The drawing which Hooke did to illustrate these notes, reproduced in plate 12, is one example of his powers as a draughtsman. On 1 March 1664/5 Pepys went again to Gresham College, where Hooke, he says, lectured about the late comet 'among other things, proving very probably that this is the very same Comet that appeared before in the year 1618 and that in such a time probably it will appear again', which, as far as I know, was the first suggestion of a periodic orbit for a comet. Again, no account. Oldenburg does, in general, seem to have done his best to discredit Hooke's work. There is no account of Hooke's book *Lampas* in the *Transactions* edited by Oldenburg. Many instances can be given of the way in which Oldenburg intrigued against Hooke. That the Society took Oldenburg's part against him on the occasion of their open dispute was a great blow to Hooke. To Hooke's precise accusations Oldenburg replied: 'The publisher of this tract intends to take another opportunity of justifying himself against the aspersions and calumnies of an immoral postscript put to a book called *Lampas*, published by R. Hooke, till which time, it is hoped, the candid reader will suspend his judgment', but never said any more. It may be noted that Newton liked Oldenburg not much better than Hooke did, in spite of the obsequiousness shown by Oldenburg to him,



Hooke's drawing of the process of felt-making. The lecture which Hooke gave on felt-making is referred to in Pepys' *Diary*.



An original Hooke microscope, now in the Science Museum. (Crown copyright. From the Court Collection in the Science Museum. Reproduced by kind permission of the Science Museum.)



Church at Willen in Buckinghamshire designed by Robert Hooke for Dr Busby.
(From a water-colour by A. E. Richardson, R.A.)

writing, for instance, 'moved by other reasons to decline, as much as Mr Oldenburg's importunity and ways to engage me in dispute will permit'.

In general, it can be shown that Hooke had much to bear from Oldenburg. The Society, too, did not treat him very well. As Curator he was always 'ordered' to 'try the experiment', 'try this by himself at home' or to 'bring in an account in writing'. Although—or because—for long periods he brought in a new experiment at practically every meeting, for little and very irregular pay, on 14 November 1670 'it was resolved that Mr Hooke be summoned to attend the next meeting of Council to receive their rebuke for the neglect of his office'.

Yes, Hooke was irritable and vain—he thought that he had performed great things in science, which is no sin, but he sometimes said so, which is. The fortunate man gets someone else to say it for him, as Francis Bacon points out in his *Essay Of Friendship*. But he was sick, very sick; an ill man day in and day out. Only occasionally, however, was he sorry for himself or too ill to work; only occasionally do we have such an entry as 'Ill all day, miserere mei deus'—God have pity on me! If he ate anything that 'agreed' he noted it in his diary. Terrible headaches, vomiting, giddiness, sleeplessness, fearful dreams all vexed him. He was perpetually administering medicaments of various kinds to himself, in the hope of relief; whether on the whole this did more harm than good, who shall say?

Yes, Hooke was irritable, especially when he thought himself slighted. He was impatient with fools. He was vexed with Sir John Cutler for promising him a stipend and then, after being much praised for his generosity, not paying it. But mark one or two strange things. Although Thomas Molyneux said that he was 'the most ill-natured, conceited man in the world, hated and dispised by most of the Royal Society'—why not by all?—he was unanimously asked to be Secretary when Oldenburg died. He was on good terms with most of the first-class intellects of the time—he dined with Evelyn and stayed for weeks at his house, was invited by Sydenham (who liked Oldenburg as little as Hooke did) to stay with him for six weeks, was on intimate terms with Christopher Wren (there is an entry of his buying a hobby horse for Wren's son), was very friendly with John Aubrey, who wrote of him in the highest terms, and was constantly in the company of Haak, Abraham Hill, Daniel Whistler and others. Busby, his schoolmaster kept up friendship with him all his life; in 1679 Hooke designed a church for him (plate 14). As for Tompion, the famous clockmaker, he occasionally enters him as 'a slug', 'a rascall', but they worked together for years, and, as supreme craftsman and supreme inventor, no doubt knew one another's worth—lost their tempers with one another and met next time as if nothing had happened.

We are told that Hooke was mean and miserly. In particular Ned Ward—'the filthy and facetious Ned Ward'—describes Hooke as a member of the Split-Farthing Club. My belief (for which I have no evidence except the characters of the two men) is that Ward, who was an innkeeper, had attempted to swindle Hooke and come off worse, for from the diary edited by Robinson and Adams it is clear that Hooke was, anyhow for the years that it covers, a sociable and free-spending man, a frequenter of coffee houses who smoked and drank with his friends. He also entertained them with wine at his lodgings and bought brandy by the gallon and wine by the dozen.

An entry like 'Drunk. Promised to pay whatever I signed to be paid' shows his usual directness, but is hard to reconcile with the miserable recluse that he is painted as by ill-wishers. He also drank tea, which cost 25 shillings a pound, at a time when a shilling was a purchasing sum. It is quite possible that Hooke was a man who did not mind spending freely on himself and his friends but disliked wasting money or being cheated—a character that always puzzles the English but not the Scots. He spent large sums on books. Yes, he was an irritable sick man, but he had his friends—and some of them were the most outstanding men of the time—and they seem to have stuck to him. And, at his funeral, anyway, the Royal Society behaved handsomely. John Ward, author of the *Lives of the Professors of Gresham College*, says well: 'Had he been more steady in his pursuits, and perfected one discovery before he entered upon another, he might perhaps in some cases have done greater service to the public, and prevented what often gave him uneasiness, the fear of losing credit of them by others, who built upon his foundation.' Always it must be remembered that he was without influence or position, other than that which he won for himself the hard way. He had no family, no profession of power, no institution behind him. It was not till 1691 that he was awarded his doctorate. In general, it seems to be just to say that while irritable, suspicious and prone to sudden anger under provocation, often severe, he bore no grudge to great men, however they treated him, and was, in large matters and in his devotion to science, a brave, and in some respects a noble, character.

You may think that, carried away by youthful enthusiasm, I have exaggerated the genius of Hooke, that, when I say that I have not even referred to half his achievements, I am using the language of political protestation. Let me, then, cite a few of the judgements of men of enduring reputation in science who have both studied Hooke's works and pronounced upon them. Thomas Young, writing of the interference of light says: 'This, I assert, is a most powerful argument in favour of the theory which I had before revived: there was nothing that could have led to it in any author with whom I am acquainted, except some imperfect hints in those inexhaustible but neglected mines of nascent inventions, the works of the great Dr Robert Hooke. . . .' Lalande, the great astronomer, referring to the *Micrographia*, *Cutlerian Lectures* and *Posthumous Works* says: 'Dans ces ouvrages on trouve les idées de la plupart de nos instruments modernes: c'était le Newton de la mécanique.' Another famous astronomer Mädler, says: 'Denn dass Hooke zu den scharfsinnigsten Denkern gehört, die jemals gelebt, kann nicht verkannt werden.' Lasswitz, who wrote the standard history of atomism, says of Hooke that he is 'eine der eigentümlichsten Erscheinungen unter den hervorragenden Forschern seiner Zeit. An allen wichtigeren Fragen der Astronomie, Mechanik und Physik beteiligt. . . macht er mit rastloser Arbeitskraft eine grosse Anzahl wertvoller Erfindungen und Entdeckungen', and devotes a chapter to his contributions to the theory of atomic vibrations. Turning to geology our own Sir Archibald Geikie, a cautious Scot, refers to him as 'one of the most brilliant, ingenious and versatile intellects of the seventeenth century', and, considering his geological work, speaks of 'his remarkable powers of acute observation and sagacious reflection'. Finally, J. F. W. Herschel calls him 'The great contemporary, and almost the worthy rival,

of Newton'. I have chosen an English physicist; a French, a German and a British astronomer; a German philosopher and a British geologist, all of first rank—I could, I assure you, without difficulty have added lesser men to the number. I think that I can claim that all those who have gone direct to Hooke have conceived the highest admiration for his astonishing industry, his whole-hearted devotion to science, his inventiveness, his ingenuity, his fertility and his brilliant theoretical insight. He rendered, with unflagging zeal, services to our Society which have not received the deserved acknowledgement. In his time the Society yielded him scant recompense, and, at any rate in his youthful prime, a grudging recognition. I have tried, while not concealing his faults, to win your sympathy for a much tried man and your approbation, your respect for his astonishing accomplishments. Admire Robert Hooke; he is worthy of your admiration.

Some theorems on perturbation theory. II

BY E. C. TITCHMARSH, F.R.S.

(Received 10 January 1950)

The theory of perturbations given in part I (Titchmarsh 1949) is extended to the so-called degenerate case.

1. In my previous paper (Titchmarsh 1949) I discussed the variation of the eigenvalues and eigenfunctions of the differential equation

$$\frac{d^2\phi}{dx^2} + \{\lambda - q(x)\}\phi = 0, \quad (1.1)$$

when the function $q(x)$ is replaced by $q(x) + \epsilon\sigma(x)$, where ϵ is a parameter which tends to zero. A brief account of the extension of this theory to the partial differential equation

$$\nabla^2\phi + \{\lambda - q(x, y)\}\phi = 0 \quad (1.2)$$

was given, and it was made clear that the method would also extend to similar equations with more than two independent variables. Throughout the paper it was assumed that to each eigenvalue there corresponded just one eigenfunction, so that the phenomenon known as 'degeneracy' did not occur.

The object of the present paper is to extend the analysis to the degenerate case. This first occurs naturally in two-dimensional problems. For example, if $q(x, y) = x^2 + y^2$ in (1.2), and the region concerned is the whole (x, y) plane, then the functions

$$(2^{m+n}m!n!\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(x^2+y^2)} H_m(x) H_n(y),$$

where $m + n = N$, all correspond to the eigenvalue $\lambda = 2N + 2$. We therefore consider as typical the two-dimensional case, though the analysis can be extended at once to the case of more than two dimensions.

The formalities of this problem are well known (see, for example, Schrödinger 1928, pp. 69–96, and Courant & Hilbert 1931, pp. 298–300). However, it may not be obvious how they should be connected with the hypotheses of the previous paper, so the required analysis will be given here.

2. We consider the equation (1.2) and the corresponding perturbed equation

$$\nabla^2 \phi + \{\lambda - q(x, y) - \epsilon \sigma(x, y)\} \phi = 0, \quad (2.1)$$

the region in each case being the whole (x, y) plane. We assume that the spectrum corresponding to the unperturbed problem is discrete, its only limit-point being $+\infty$. For simplicity in writing, we shall suppose that the multiple eigenvalue considered is the smallest, λ_0 . Let it be k -fold, i.e. suppose that there are k normal orthogonal eigenfunctions $\psi_0(x, y), \dots, \psi_{k-1}(x, y)$ all corresponding to the eigenvalue λ_0 . The next greater eigenvalue will be denoted by λ_k , so that $\lambda_1, \dots, \lambda_{k-1}$, if they occur in the formulae, are all equal to λ_0 . λ_k and all subsequent eigenvalues will be treated as simple, though this does not make any material difference to the analysis.

We also assume that

$$\int_0^\infty \int_0^{2\pi} \left(\frac{\partial \psi_n}{\partial r} \right)^2 r dr d\theta$$

is convergent for each n . As was shown in the previous paper, this is certainly true if $q(x, y) \geq 0$.

With regard to the perturbed equation (2.1), we assume that it also has discrete eigenvalues Λ_n , their only limit-point being $+\infty$, with eigenfunctions $\Psi_n(x, y)$. We shall consider only the simplest case, in which $\epsilon > 0$ and $\sigma(x, y) \geq 0$, and it will be assumed that

$$b_n = \int_{-\infty}^\infty \int_{-\infty}^\infty \{\sigma(x, y) \psi_n(x, y)\}^2 dx dy$$

is finite for every n .

As in the case of one variable, each Λ_n decreases steadily as $\epsilon \rightarrow 0$, while remaining greater than λ_0 . Hence each Λ_n tends to a limit as $\epsilon \rightarrow 0$. An argument similar to that of the previous paper shows that $\Lambda_0 \rightarrow \lambda_0$, but the behaviour of Λ_n for $n > 0$ is not immediately clear.

3. Let

$$a_{m,n} = \int_{-\infty}^\infty \int_{-\infty}^\infty \sigma(x, y) \psi_m(x, y) \psi_n(x, y) dx dy, \quad (3.1)$$

$$A_{m,n} = \int_{-\infty}^\infty \int_{-\infty}^\infty \sigma(x, y) \Psi_m(x, y) \psi_n(x, y) dx dy, \quad (3.2)$$

$$D_{m,n} = \int_{-\infty}^\infty \int_{-\infty}^\infty \Psi_m(x, y) \psi_n(x, y) dx dy. \quad (3.3)$$

It follows from our hypotheses that these integrals are all convergent. As in the previous paper, we have

$$(\Lambda_m - \lambda_n) D_{m,n} = \epsilon A_{m,n} \quad (3.4)$$

for all values of m and n .

Now for $m = k, k+1, \dots$ we have

$$\Lambda_m - \lambda_0 \geq \lambda_k - \lambda_0 = \delta$$

say, where δ is positive. Hence

$$|D_{m,n}| \leq \frac{\epsilon}{\delta} |A_{m,n}| \quad (m \geq k, n < k).$$

$$\text{Hence } \sum_{m=k}^{\infty} D_{m,n}^2 \leq \frac{\epsilon^2}{\delta^2} \sum_{m=k}^{\infty} A_{m,n}^2 \leq \frac{\epsilon^2}{\delta^2} \sum_{m=0}^{\infty} A_{m,n}^2 = \frac{\epsilon^2}{\delta^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{\sigma(x, y) \psi_n(x, y)\}^2 dx dy,$$

by the Parseval theorem for the perturbed equation. Thus, for any given n less than k ,

$$\sum_{m=k}^{\infty} D_{m,n}^2 = O(\epsilon^2).$$

The Parseval theorem for the perturbed equation also gives

$$\sum_{m=0}^{\infty} D_{m,n}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_n^2(x, y) dx dy = 1.$$

$$\text{Hence } \sum_{m=0}^{k-1} D_{m,n}^2 = 1 + O(\epsilon^2) \quad (n < k). \quad (3.5)$$

Now the Parseval theorem for the unperturbed equation gives

$$\sum_{n=0}^{\infty} D_{m,n}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi_m^2(x, y) dx dy = 1. \quad (3.6)$$

$$\text{Hence } \sum_{m=0}^{k-1} \sum_{n=0}^{\infty} D_{m,n}^2 = k. \quad (3.7)$$

Subtracting each of the equations (3.5) from (3.7), it follows that

$$\sum_{m=0}^{k-1} \sum_{n=k}^{\infty} D_{m,n}^2 = O(\epsilon^2).$$

$$\text{Hence } \sum_{n=k}^{\infty} D_{m,n}^2 = O(\epsilon^2) \quad (m < k). \quad (3.8)$$

From (3.6) and (3.8) it follows that

$$\sum_{n=0}^{k-1} D_{m,n}^2 = 1 + O(\epsilon^2) \quad (m < k). \quad (3.9)$$

We now use the Parseval formula for the unperturbed equation in the form

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) g(x, y) dx dy = \sum_{n=0}^{\infty} c_n d_n,$$

where the c_n are the 'Fourier coefficients' of $f(x, y)$, and the d_n those of $g(x, y)$. Taking

$$f(x, y) = \Psi_{\mu}(x, y), \quad g(x, y) = \sigma(x, y) \psi_{\nu}(x, y),$$

$$\text{it follows that } A_{\mu,\nu} = \sum_{n=0}^{\infty} D_{\mu,n} a_{\nu,n}. \quad (3.10)$$

Also for $\mu = 0, 1, \dots, k-1$, and any given ν ,

$$\begin{aligned} \left| \sum_{n=k}^{\infty} D_{\mu,n} a_{\nu,n} \right| &\leq \left(\sum_{n=k}^{\infty} D_{\mu,n}^2 \sum_{n=k}^{\infty} a_{\nu,n}^2 \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{n=k}^{\infty} D_{\mu,n}^2 \sum_{n=0}^{\infty} a_{\nu,n}^2 \right)^{\frac{1}{2}} \\ &= \left\{ \sum_{n=k}^{\infty} D_{\mu,n}^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{\sigma(x, y) \psi_{\nu}(x, y)\}^2 dx dy \right\}^{\frac{1}{2}} = O(\epsilon) \end{aligned}$$

$$\text{by (3.8). Hence } A_{\mu,\nu} = \sum_{n=0}^{k-1} D_{\mu,n} a_{\nu,n} + O(\epsilon) \quad (\mu < k). \quad (3.11)$$

From (3.4) and (3.11) we obtain

$$(\Lambda_\mu - \lambda_0) D_{\mu,\nu} = \epsilon \sum_{n=0}^{k-1} D_{\mu,n} a_{\nu,n} + O(\epsilon^2) \quad (\mu < k). \quad (3.12)$$

Writing $X_\mu = (\Lambda_\mu - \lambda_0)/\epsilon$, (3.12) takes the form

$$X_\mu D_{\mu,\nu} = \sum_{n=0}^{k-1} D_{\mu,n} a_{\nu,n} + O(\epsilon). \quad (3.13)$$

If we could ignore the terms $O(\epsilon)$ and eliminate the numbers $D_{\mu,\nu}$ in the usual way, it would follow that the X_μ would approximate to the roots of the equation

$$\begin{vmatrix} a_{0,0} - x & a_{0,1} & \cdots & \cdots \\ a_{1,0} & a_{1,1} - x & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & a_{k-1,k-1} - x \end{vmatrix} = 0.$$

Since the numbers $D_{\mu,\nu}$ also depend on ϵ , it is not immediately clear that this process is valid.

4. We now observe that the choice of the k eigenfunctions associated with a k -fold eigenvalue is not unique. If one such set is $\psi_\mu(x, y)$ and $b_{\mu,\nu}$ are the coefficients in an orthogonal transformation, then the functions $\psi'_\nu(x, y)$ defined by

$$\psi'_\nu(x, y) = \sum_{\mu=0}^{k-1} b_{\mu,\nu} \psi_\mu(x, y) \quad (4.1)$$

is another set with the same properties, as is easily verified. Now a familiar theorem of algebra asserts that a quadratic form

$$A = \sum_{\mu=0}^{k-1} \sum_{\nu=0}^{k-1} a_{\mu,\nu} \xi_\mu \xi_\nu \quad (4.2)$$

can be reduced by an orthogonal transformation

$$\xi_\mu = \sum_{\rho=0}^{k-1} b_{\mu,\rho} \xi'_\rho \quad (4.3)$$

to the form

$$A = \sum_{\mu=0}^{k-1} \sum_{\nu=0}^{k-1} a'_{\mu,\nu} \xi'_\mu \xi'_\nu, \quad (4.4)$$

where $a'_{\mu,\nu} = 0$ if $\mu \neq \nu$. If the coefficients $a_{\mu,\nu}$ are defined by (3.1), then

$$\begin{aligned} A &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma(x, y) \sum_{\mu=0}^{k-1} \xi_\mu \psi_\mu(x, y) \sum_{\nu=0}^{k-1} \xi_\nu \psi_\nu(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma(x, y) \sum_{\mu=0}^{k-1} \xi'_\mu \psi'_\mu(x, y) \sum_{\nu=0}^{k-1} \xi'_\nu \psi'_\nu(x, y) dx dy \end{aligned} \quad (4.5)$$

by (4.1) and (4.3). Comparing (4.4) and (4.5),

$$a'_{\mu,\nu} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma(x, y) \psi'_\mu(x, y) \psi'_\nu(x, y) dx dy.$$

In the transformed system we therefore have $a'_{\mu,\nu} = 0$ for $\mu \neq \nu$.

5. To save rewriting the results of § 3, we shall suppose that the transformation of § 4 has previously been effected, so that the numbers $a_{\mu,\nu}$ have the property that $a_{\mu,\nu} = 0$ if $\mu < k$, $\nu < k$, $\mu \neq \nu$. The equations (3.13) then take the form

$$(X_\mu - a_{\nu,\nu}) D_{\mu,\nu} = O(\epsilon) \quad (\mu < k, \nu < k). \quad (5.1)$$

To take the simplest case, suppose that the numbers $a_{0,0}, \dots, a_{k-1,k-1}$ as now defined are all different. Since the order in which they are numbered is so far arbitrary, we can suppose that

$$a_{0,0} < a_{1,1} < \dots < a_{k-1,k-1}.$$

Consider the equations (5.1) with a fixed μ . By (3.9),

$$\sum_{\nu=0}^{k-1} D_{\mu,\nu}^2 \geq \frac{1}{2},$$

if ϵ is small enough. Hence

$$D_{\mu,\nu}^2 \geq \frac{1}{2k}$$

for some value of ν , depending on μ . Hence (5.1) gives

$$X_\mu - a_{\nu,\nu} = O(\epsilon) \quad (5.2)$$

for this ν .

If two different numbers ν and ν' corresponded in this way to the same number μ , we should have

$$a_{\nu,\nu} - a_{\nu',\nu'} = O(\epsilon),$$

contrary to the hypothesis that the $a_{\nu,\nu}$ are all unequal. Hence just one ν corresponds to each μ . Also (5.2) cannot hold with two different values of μ and the same ν . For suppose that (5.2) holds with $\nu = \nu'$ and $\mu = \mu'$ and $\mu = \mu''$ ($\mu' \neq \mu''$). Then

$$\begin{aligned} X_\mu - a_{\nu,\nu} &= X_\mu - a_{\nu',\nu'} + a_{\nu',\nu'} - a_{\nu,\nu} \\ &= a_{\nu',\nu'} - a_{\nu,\nu} + O(\epsilon) \end{aligned}$$

for $\mu = \mu'$ and $\mu = \mu''$. Hence by (5.1)

$$D_{\mu,\nu} = O(\epsilon) \quad (\mu = \mu', \mu = \mu'', \nu \neq \nu').$$

Hence by (3.9)

$$D_{\mu',\nu'}^2 = 1 + O(\epsilon^2), \quad D_{\mu'',\nu'}^2 = 1 + O(\epsilon^2).$$

Hence

$$\sum_{m=0}^{k-1} D_{m,\nu'}^2 \geq 2 + O(\epsilon^2),$$

which is inconsistent with (3.5).

There is therefore a one-one correspondence between the numbers μ and ν for which (5.2) holds. If the numbers Λ_μ , and so also X_μ , are numbered in increasing order of magnitude, it follows that

$$X_\nu - a_{\nu,\nu} = O(\epsilon) \quad (\nu < k).$$

Hence

$$\Lambda_\nu = \lambda_0 + \epsilon a_{\nu,\nu} + O(\epsilon^2) \quad (\nu < k). \quad (5.3)$$

This is the required extension of theorem 4 of the previous paper.

6. We next want to prove that $\Psi_n(x, y) \rightarrow \psi_n(x, y)$ for $n < k$. Now it follows from (5.1) and (5.3) that

$$D_{\mu,\nu} = O(\epsilon) \quad (\mu < k, \nu < k, \mu \neq \nu).$$

Hence by (3.5)
$$D_{n,n} = 1 + O(\epsilon^2) \quad (n < k). \quad (6.1)$$

Hence
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{\Psi_n(x, y) - \psi_n(x, y)\}^2 dx dy = 2 - 2D_{n,n} = O(\epsilon^2). \quad (6.2)$$

Also (Courant & Hilbert 1937, p. 250 (7))

$$\psi_n(x_0, y_0) = \frac{1}{\pi R^2} \iint_{r \leq R} \psi_n r dr d\theta + \frac{1}{2\pi} \iint_{r \leq R} \left\{ \frac{1}{2} \left(1 - \frac{r^2}{R^2} \right) - \log \frac{R}{r} \right\} (q - \lambda_n) \psi_n r dr d\theta,$$

where on the right-hand side $\psi_n = \psi_n[r, \theta] = \psi_n(x, y)$, etc., the integration being over a circle with centre (x_0, y_0) and radius R . Similarly,

$$\Psi_n(x_0, y_0) = \frac{1}{\pi R^2} \iint_{r \leq R} \Psi_n r dr d\theta + \frac{1}{2\pi} \iint_{r \leq R} \left\{ \frac{1}{2} \left(1 - \frac{r^2}{R^2} \right) - \log \frac{R}{r} \right\} (q + \epsilon\sigma - \Lambda_n) \Psi_n r dr d\theta.$$

Hence

$$\begin{aligned} \Psi_n(x_0, y_0) - \psi_n(x_0, y_0) &= \frac{1}{\pi R^2} \iint_{r \leq R} (\Psi_n - \psi_n) r dr d\theta \\ &\quad + \frac{1}{2\pi} \iint_{r \leq R} \left\{ \frac{1}{2} \left(1 - \frac{r^2}{R^2} \right) - \log \frac{R}{r} \right\} (q - \lambda_0) (\Psi_n - \psi_n) r dr d\theta \\ &\quad + \frac{1}{2\pi} \iint_{r \leq R} \left\{ \frac{1}{2} \left(1 - \frac{r^2}{R^2} \right) - \log \frac{R}{r} \right\} (\epsilon\sigma + \lambda_0 - \Lambda_n) \Psi_n r dr d\theta. \end{aligned}$$

Using the Schwarz inequality and (6.2), it follows that the first two terms on the right-hand side are $O(\epsilon)$; and similarly so is the third, since

$$\epsilon\sigma + \lambda_0 - \Lambda_n = O(\epsilon).$$

Hence
$$\Psi_n(x_0, y_0) - \psi_n(x_0, y_0) = O(\epsilon), \quad (6.3)$$

and clearly this is uniform over any finite region.

7. As in the previous paper, we can make further approximations to Λ_n and $\Psi_n(x, y)$ if we make further assumptions. As before, we have

$$D_{n,m} = \frac{\epsilon a_{m,n}}{\Lambda_n - \lambda_m} + O\left(\frac{\epsilon^2 b_m^\dagger}{\lambda_m}\right) \quad (m \geq k, n < k), \quad (7.1)$$

where b_m is defined in § 2. By (3.4) and (3.10)

$$\begin{aligned} (\Lambda_n - \lambda_0) D_{n,\nu} &= \epsilon \sum_{m=0}^{\infty} D_{n,m} a_{m,\nu} \\ &= \epsilon D_{n,\nu} a_{\nu,\nu} + \epsilon \sum_{m=k}^{\infty} D_{n,m} a_{m,\nu}, \end{aligned} \quad (7.2)$$

since now $a_{m,\nu} = 0$ ($\nu < k, m \neq \nu$). Taking $\nu = n$ and proceeding as before, we therefore obtain

$$\Lambda_n = \lambda_0 + \epsilon a_{n,n} + \epsilon^2 \sum_{m=k}^{\infty} \frac{a_{m,n}^2}{\lambda_n - \lambda_m} + O(\epsilon^3), \quad (7.3)$$

provided that

$$\sum_m \left| \frac{a_{m,n} b_m^\dagger}{\lambda_m} \right| \quad (7.4)$$

is convergent.

As regards $\Psi_n(x, y)$ we have, as before,

$$\Psi_n(x, y) = \sum_{\nu=0}^{\infty} D_{n,\nu} \psi_{\nu}(x, y). \quad (7.5)$$

Now if $n < k$ and $\nu < k$, (7.1) and (7.2) give

$$(\Lambda_n - \lambda_0) D_{n,\nu} = \epsilon D_{n,\nu} a_{\nu,\nu} + \epsilon^2 \sum_{m=k}^{\infty} \frac{a_{m,n} a_{m,\nu}}{\Lambda_n - \lambda_m} + O(\epsilon^3),$$

provided that (7.4) is convergent; and we can clearly replace Λ_n by λ_n with error $O(\epsilon^3)$. Hence by (5.3)

$$\{\epsilon a_{n,n} - \epsilon a_{\nu,\nu} + O(\epsilon^2)\} D_{n,\nu} = \epsilon^2 \sum_{m=k}^{\infty} \frac{a_{m,n} a_{m,\nu}}{\lambda_n - \lambda_m} + O(\epsilon^3).$$

Hence if $\nu < k$, $n < k$,

$$D_{n,\nu} = \frac{\epsilon}{a_{n,n} - a_{\nu,\nu}} \sum_{m=k}^{\infty} \frac{a_{m,n} a_{m,\nu}}{\lambda_n - \lambda_m} + O(\epsilon^2). \quad (7.6)$$

From (7.5), (6.1) and (7.6) (7.1), it follows that

$$\Psi_n(x, y) = \psi_n(x, y) + \epsilon \sum_{\substack{\nu < k \\ \nu \neq n}} \frac{\psi_{\nu}(x, y)}{a_{n,n} - a_{\nu,\nu}} \sum_{m=k}^{\infty} \frac{a_{m,n} a_{m,\nu}}{\lambda_n - \lambda_m} + \epsilon \sum_{\nu=k}^{\infty} \frac{a_{n,\nu} \psi_{\nu}(x, y)}{\lambda_n - \lambda_{\nu}} + O(\epsilon^2), \quad (7.7)$$

provided that (7.4) and

$$\sum \left| \frac{b_{\nu}^{\frac{1}{2}} \psi_{\nu}(x, y)}{\lambda_{\nu}} \right| \quad (7.8)$$

are convergent.

REFERENCES

- Courant, R. & Hilbert, D. 1931 *Methoden der mathematischen Physik*, 1 (zw. Aufl.). Berlin: Springer.
 Courant, R. & Hilbert, D. 1937 *Methoden der mathematischen Physik*, 2. Berlin: Springer.
 Schrödinger, E. 1928 *Collected papers on wave mechanics*. London and Glasgow: Blackie.
 Titchmarsh, E. C. 1949 *Proc. Roy. Soc. A*, **200**, 34-46.

Studies on the ionization produced by metallic salts in flames

I. The determination of the collision frequency of electrons in coal-gas/air flames

BY H. BELCHER AND T. M. SUGDEN

Department of Physical Chemistry, University of Cambridge

(Communicated by R. G. W. Norrish, F.R.S.—Received 6 October 1949)

As an introduction to the study of reactions contingent on ionization in flames, an experimental measurement has been made of the collision frequency of electrons with molecules in coal-gas/air flames, containing added alkali metal salt. This quantity is an important parameter in the expression relating the electron content of a flame with the attenuation of centimetric radio waves by it. This attenuation has been chosen as a convenient method of investigating flame ionization. The form of the results obtained agree well with the predictions of theory, a uniform difference of about 20 % between measured collision frequency and that calculated on a very simple gas kinetic hypothesis being obtained. A suitable conversion factor has been evolved for proceeding from attenuation of 3 cm. waves to electron concentration/cm.³.

1. INTRODUCTORY

There have been several previous attempts to elucidate the extent of the ionization produced when a salt of an alkali metal is introduced into a flame, which have not been entirely satisfactory from two main points of view. On the experimental side, it has been the practice to study the conductivity of the flames using d.c. or low-frequency a.c.; these result in space-charge effects around the electrodes, which are difficult to eliminate or to compensate for. In addition, doubt with regard to the mobility of the free electrons, which give rise to the greater part of the current in the flame, renders calculation of the actual number of collisions open to error. This is important, since in general, the electrical conductivity depends on the number of free electrons, the rate at which they collide with molecules and the frequency of the impressed field. The position in the sphere of conductivity measurements has been well summarized in a review by Wilson (1944).

It was decided to study the electron concentration in flames by measuring the attenuation of centimetric waves passing through them, and it will be shown in this first paper that, under the experimental conditions employed, this attenuation is directly proportional to the electron concentration, and that the conversion factor may be established to a fair degree of accuracy, either theoretically or by experiment. The method has the advantage of leaving the flame undisturbed and has previously been used by Andrew, Axford & Sugden (1948) in a crude form in the case of a transient flame, where it was shown to give results compatible with those from the d.c. conductivity method. In order to establish a value for the conversion factor from attenuation to electron concentration it is necessary to measure the collision frequency of electrons with molecules in the flames, and these basic measurements will be described in this paper.

The general purpose of the investigations is that of attempting to correlate ionization in flames with flame temperature, with reference to all side reactions which may affect the ion content, and in particular to find whether any evidence of temperature disequilibrium of reactions is shown, and to what extent.

2. DETERMINATION OF THE ELECTRON CONCENTRATION

(a) Simple theoretical considerations

The passage of electromagnetic radiation through an isotropic medium containing n free ions/cm.³ is accompanied by an attenuation of the waves, since the free ions are set into sympathetic vibration and lose such directed momentum at encounters with large molecules. If x is the displacement of such an ion, of charge e and mass m , it is possible to represent its motion by the equation

$$\ddot{x} + \omega_1 \dot{x} = \frac{eE}{m}, \quad (1)$$

where E is the electric field and ω_1 the collision frequency of the ion with molecules, i.e. the effect of discontinuous molecular collisions on the induced vibratory motion is represented by a viscous retarding force proportional to the velocity of the ion. This involves a number of assumptions, notably that the electric field is not sufficiently strong to affect the collision frequency of the ions, and that several vibrations may occur along one linear mean free path of an ion. The former was ensured in all the present work by working at a sufficiently low power but the latter assumption is not justified. It implies an erroneous averaging of the frequencies used, and to obtain an exact solution it is necessary to use an expression for the distribution functions of velocities, which has been done by Margenau (1945), to whose work reference will be made later. For the moment, the above equation (1) will be considered, however, on account of the simplicity of results to which it leads. The application of the above equation in classical electrodynamics is simple and has been carried out previously several times, leading to the following expressions for the effective dielectric constant and electrical conductivity:

$$\epsilon = 1 - \frac{\omega_c^2}{\omega^2 + \omega_1^2}, \quad \sigma = \frac{\omega_c^2 \omega_1}{4\pi(\omega^2 + \omega_1^2)}, \quad (2)$$

if $\epsilon = 1$ and $\sigma = 0$ in the absence of ions (Gaussian units will be used), where ω is the angular frequency of the radiation ($= 2\pi f$ if f is the frequency) and $\omega_c^2 = \frac{4\pi n e^2}{m}$. It is sometimes more useful to write these in terms of real and imaginary conductivities σ_r and σ_i defined by

$$\sigma_r = \sigma, \quad \sigma_i = \frac{\omega}{4\pi} (1 - \epsilon). \quad (2a)$$

The coefficient of attenuation χ , where $e^{-\chi r}$ determines the field transmission in the direction r , is then given by

$$\chi^2 = \frac{\omega^2 \mu}{2c_0^2} \left[\sqrt{\epsilon^2 + \left(\frac{4\pi\sigma}{\omega} \right)^2} - \epsilon \right], \quad (3)$$

where μ = magnetic permeability and c_0 is the velocity of light *in vacuo*. In a case where ϵ is not sensibly different from unity and $\left(\frac{4\pi\sigma}{\omega}\right)^2 \ll 1$, a simple binomial expansion may be performed, leading to

$$\chi = \frac{2\pi}{c_0} \sigma \sqrt{\mu} = \frac{2\pi e^2 n}{mc_0} \left(\frac{\omega_1}{\omega^2 + \omega_1} \right) \sqrt{\mu}. \quad (4)$$

It is useful at this point to consider probable values of ϵ and σ in the systems under investigation. ω ranges from about 2×10^{10} (10 cm. wave-length) to about 3×10^{11} . A first approximation to ω_1 may be obtained from the kinetic theory of gases by considering point electrons, and molecules with collision diameters derived from viscosity measurements; then

$$\text{collision frequency of an ion with molecules} = \omega_1 = pd^2 \sqrt{\left(\frac{\pi}{2mkT} \right)}, \quad (5)$$

where p is the pressure, d the molecular diameter and m the electronic mass. It should be noted that electrons will be much more effective in attenuations than any large ions of the same charge since $\omega^2 \propto 1/m$. Inserting as reasonable values

$$p = 10^6 \text{ dynes/sq.cm.}, \quad T = 2000^\circ \text{ K} \quad \text{and} \quad d = 3 \times 10^{-8} \text{ cm.},$$

a value of $\omega_1 = 7 \times 10^{10}$ is obtained. This is unlikely to be far removed from a lower limit. Numerical tests show that (4) is a reasonable approximation for χ for all the wave-lengths at $\omega_1 = 5 \times 10^{10}$ up to $n = 5 \times 10^{11}$, and up to higher electron concentrations for larger values of ω_1 . Expressing the power loss as an attenuation β in db./cm., we have $\beta = 8.7\chi$, it is therefore possible to say that at all the wave-lengths used, the simplified form (4) may be used for values of β up to 2 db./cm. attenuation. Values of $e = 4.80 \times 10^{-10}$ e.s.u., $m = 9.1 \times 10^{-28}$ g., and $c_0 = 3.00 \times 10^{10}$ cm./sec. have been used, and $\mu = 1.00$, which is quite reasonable for flame gases.

In order to test the validity of (4), a series of observations have been made of attenuation in flames at various wave-lengths; the expression may be more simply written

$$\beta = \frac{a\omega_1}{\omega^2 + \omega_1^2}, \quad (6)$$

where $a = 8.7 \left(\frac{2\pi e^2}{mc_0} \right) n$ and is constant for a given flame, or

$$\frac{1}{\beta} = \frac{\omega^2}{a\omega_1} + \frac{\omega_1}{a}. \quad (7)$$

Hence a graph of $1/\beta$ against ω^2 should be a straight line from which ω_1 and a should be calculable. The ratio of attenuations produced by two different wave-lengths should be the same for a given flame provided its electron content is such that it does not exceed the value allowing of expansion of (3).

(b) *Experimental determination of ω_1*

A suitably attenuating flame for comparison at various frequencies was obtained by burning pre-mixed coal gas and air, in known composition, in a Mèker-type burner, shown at *B* in figure 1 *a* to *d*. The burner *B* was made of an insulating refractory and had small expansion chambers at either end to give a uniform flame along its length of 18 cm.; it was about 1.5 cm. in thickness. The height of the flame varied

with the composition and rate of flow, ranging from 5 to 20 cm. The general technique adopted in each case was the same in principle, although slight variants were adopted at different frequencies. The flame was made attenuating by adding a fine spray of potassium bicarbonate solution from an atomizer to the air stream before mixing. Square-wave modulated power from an oscillator was fed from a rectangular wave-guide system into another system across a gap containing the flame (figure 1). The transmitted signal was rectified by a crystal and passed via a calibrated resistive attenuator into an amplifier limiter system shown in figure 2.

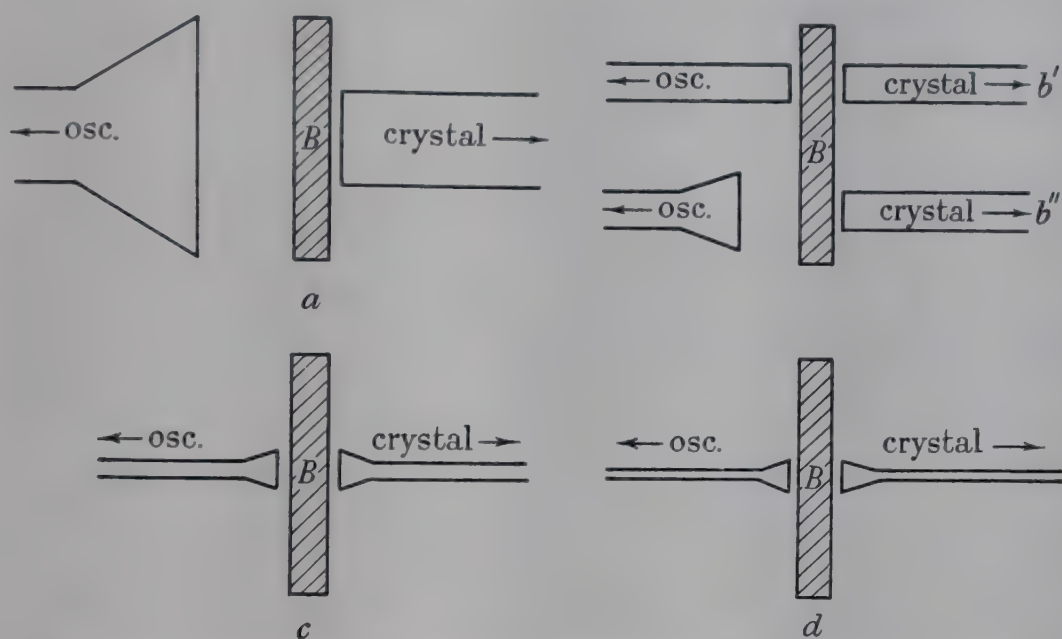


FIGURE 1. Diagrammatic wave-guide dispositions for various frequencies.

This consisted of two stages of conventional resistance-capacity coupled amplification (V_1 and V_2), giving an output which was a slightly modified form of the original square wave (see anode V_2). This was restored to d.c. by one-half of V_3 and fed to the anode of the second half, which formed part of a limiter circuit passing only the positive peaks of the signal. The height of these peaks was determined by the setting of the limiter control which varied the cathode bias on V_3 , so that the signal as shown was fed on to a final stage of amplification V_4 , after which it was recorded on a cathode-ray tube after d.c. resotation by V_5 . With this system of limiting careful observation allowed of an accuracy of ± 0.01 db. in measuring attenuations. The method of working was to set the limiting potentiometer, with the attenuating flame present, so that a definite deflexion was obtained on the oscillograph; the source of attenuation was then removed and the calibrated attenuator adjusted to bring in an attenuation giving the same deflexion. The required attenuation was thus read off directly. In the case of the 10 cm., 3 cm. and 8 mm. radiation the flame in the absence of added salt was found to be non-attenuating; at 1.2 cm., however, a small but measurable attenuation was recorded, under these conditions, amounting to 5 to 10 % of the attenuation with salt added. This is doubtless due to the high content of water vapour, which is known to have an absorption band in this region, in the flame gases.

The principal difficulty was that of obtaining a disposition of wave guides such that the same region of the flame height was studied at the various wave-lengths. Suitable dispositions were obtained by trial, the effective region of observation being found by interposition of pieces of conductor or dielectric at various points around the gap in the wave guides. The methods used are shown in figure 1; at 10 cm. (1*a*) the power was fed from a large sectoral horn well away from the flame and was picked up by a plain-ended guide placed close behind it. At 3 cm. (1*b*) two dispositions were used, both giving the same attenuation—*b'* consisting of two plain guides close to the flame and *b''* of an arrangement similar to 1*a* with a square aperture horn and a plain

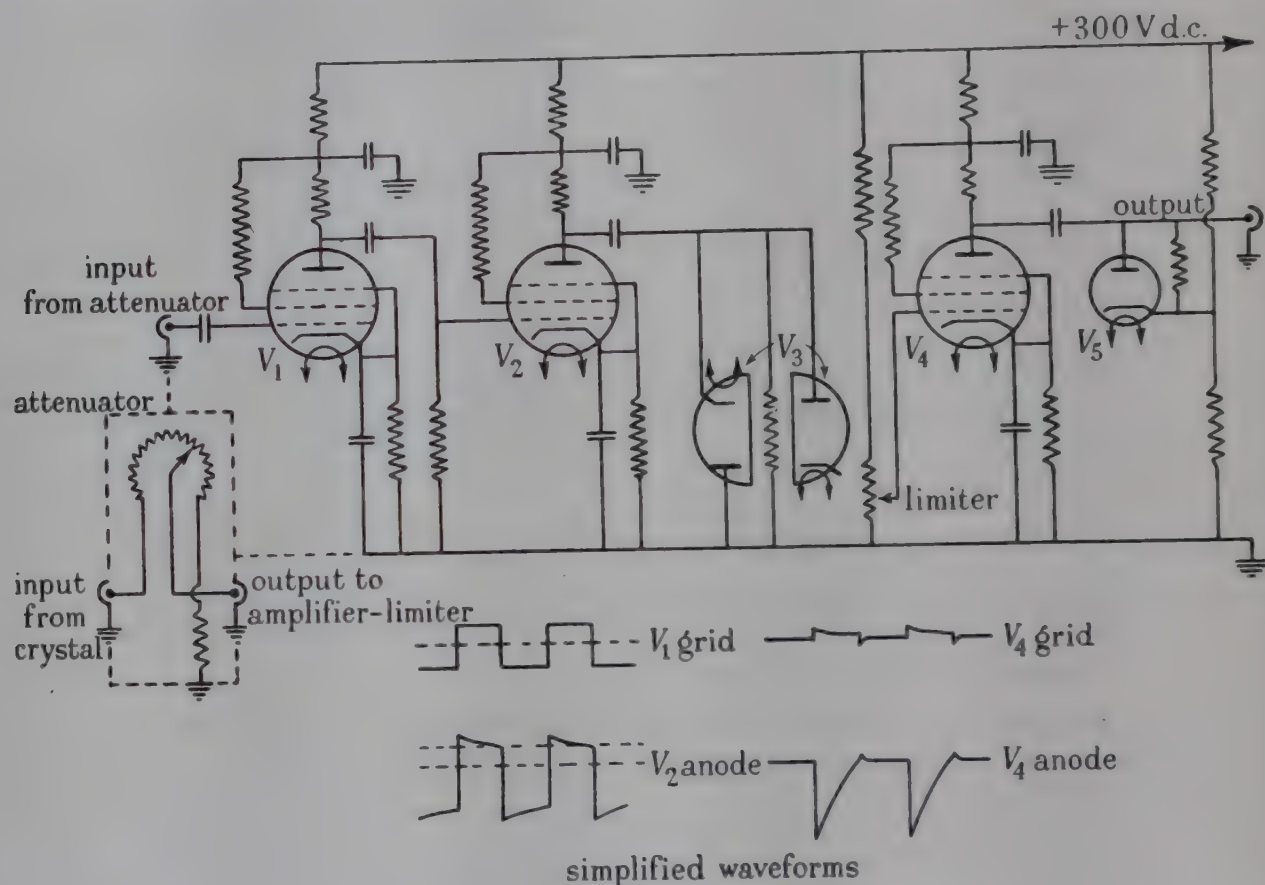


FIGURE 2. Circuit of amplifier-limiter system.

guide close to the flame. This wave-length proves to be the most useful for the laboratory flames since it combines high attenuations and hence high accuracy with a reasonably large flame; most of the systematic observations on flames have been carried out with its aid. At 1.2 cm. and 8 mm. square aperture horns were used at the ends of the guides. The region of flame studied was that immediately above the small blue cones in which a minimum of cooling together with some degree of chemical equilibrium may be expected. No difficulty was present in obtaining readings with a wide variety of flames except at 10 cm. wave-length, where only the larger flames gave good readings, the smaller ones requiring thickness correction since it was not possible to obtain readings over quite the same height of flame as in the other cases.

Low-power Klystron oscillators were used at 10, 3 and 1.2 cm., the modulation in the first two cases being obtained by rotation of a brass sector disk in a subsidiary

small gap in the guides, and in the last case by applying a square-wave modulation to the grid of the oscillator. At 8 mm. a pulsed magnetron supply was used but suitable attenuation introduced to bring the signal level down to the order of that used at the other wave-lengths. In all cases considerable attenuation (several db.) was introduced on either side of the flame gap so as to reduce the standing wave ratio in the flame gap as much as possible. No trouble was experienced with heating and cooling of the guide system since measurements could be taken rapidly. At 3, 1.2 and 0.8 cm. there was no detectable reflexion from the flame at normal incidence, but at 10 cm. some slight trouble was experienced which caused the attenuation to vary over a range of 10 % with slight variation of the distance of the burner from the pick-up guide. This may be due to reflexion at the flame surface, which is expected to be just appreciable at 10 cm. wave-length and too small to measure at lower ones, or it may arise once again from the difficulty of confining the region of observation to a sufficiently small size at the large wave-length. As mentioned previously, a small effect was observed in the absence of added salt at 1.2 cm. only, and in assessing the attenuation caused by ions, this amount, measured with pure water instead of salt in the atomizer, was subtracted from the total attenuation obtained with the salt solution.

The comparative results are very striking. The ratio of attenuations due to electrons obtained at pairs of different wave-lengths remains constant within 5 % for any given pair over a wide range of flame-gas compositions from gas rich through stoichiometric to air rich and is independent of the concentration of potassium salt added over a concentration range of ten-fold. The attenuations β ranged up to 1.5 db., which is within the range where the simple binomial expansion of the theory is valid. The concentration of salt added expressed as a fraction of the total pressure in the flame is of the order 10^{-5} of the whole, and it is therefore not expected that varying the substance added will affect this ratio. Addition of halide salts instead of bicarbonate shows this to be true, the presence of such small quantities of halogen being insufficient to cause a sensible alteration of the mean free path of the electrons.

The observed ratios are as follows—3 cm. radiation being taken as a basis, since it has been used for a much wider variety of flame measurements:

	ratio of attenuations
0.8 : 3 cm.	0.20 : 1.00
1.2 : 3 cm.	0.40 : 1.00
10 : 3 cm.	1.40 : 1.00 (for large flames only)

To some extent, therefore, the simple theory seems justified; the variation in the ratios observed at 0.8 and 1.2 cm. of about 5 % is within the experimental error since the attenuations at these wave-lengths were so small. A quantitative test of the results may be made by plotting $1/\beta$ for any given flame against ω^2 (viz. equation (7)). This has been done for a typical flame (air/gas = 2.67, where stoichiometric = 4.00, thickness 1.45 cm., temperature 2200° K as determined by the sodium line-reversal method) in figure 3, and yields an excellent straight line. The wave-lengths were accurately determined by absorption wavemeters at 10.4, 3.0, 1.27 and 0.78 cm.

From (7)
$$\frac{\text{intercept on } 1/\beta \text{ axis}}{\text{slope}} = \frac{\omega_1}{a} a \omega_1 = \omega_1^2$$

$$\frac{1}{(\text{slope}) \times (\text{intercept on } 1/\beta \text{ axis})} = \frac{a}{\omega_1} a \omega_1 = a^2$$

and using figure 2, $\omega_1 = 8.8 \times 10^{10}$ impacts per second and

$$n = a \left(\frac{mc_0}{17.4\pi e^2} \right) = 2.0 \times 10^{11} \text{ electrons/cm.}^3.$$

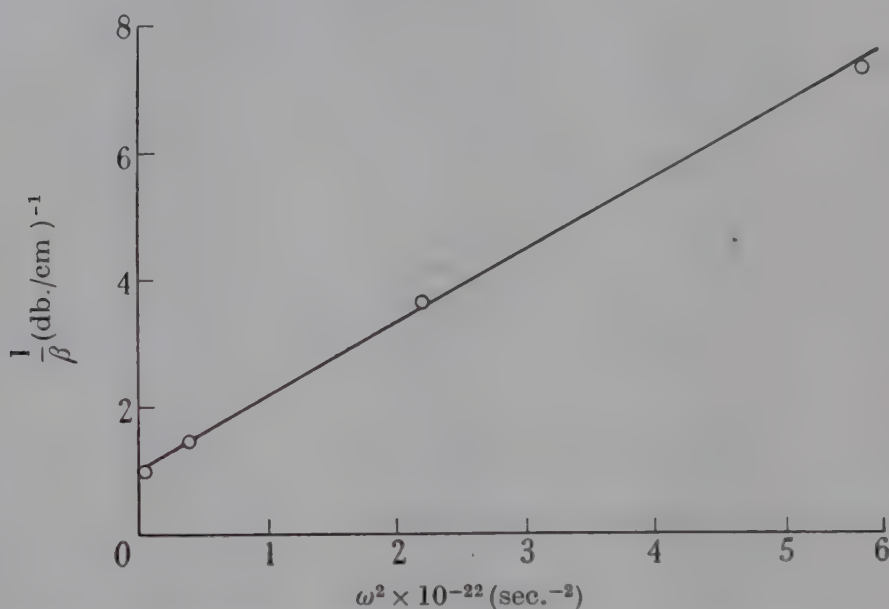


FIGURE 3. Graph of $1/\text{attenuation}$ against $(\text{frequency})^2$. Slope = $1/a\omega_1$; intercept = ω_1/a .

The first of these results is independent of the thickness measurement, and will obviously be the same for all the flames on account of the constant ratios noted above. It is only 20 % higher than the simple result obtained by treating the electron as a point and using viscosity data for the flame gases (from viscosities measured at room temperature) to derive collision diameters. The mean diameter is not expected to vary much over the range of gas compositions used, since the principal constituents (H_2 , H_2O , CO , CO_2 , N_2) have collision diameters not very different from one another. The experimental results, therefore, appear to afford a very sound vindication of the simple theory outlined above, and also provide a basis for the determination of ω_1 , which may be compared with any values which may become available from quantum-mechanical calculations.

(c) Summary of more detailed theory

It is still necessary to examine a few more detailed theoretical considerations since despite the agreement obtained above, there is no doubt that in the simple cases there are present unjustified assumptions in the averaging of velocities of electrons, since the collision frequency is very much the same as the angular frequency of the radiation used. This has been done fairly thoroughly by Margenau (1945) who replaces the differential equation of motion (1) by a similar one for the distribution function of velocities. In the simplest case, where the intensity of the radiation is

insufficient to change the number of collisions, as in the present case, and where a Maxwellian distribution of velocities is assumed, this leads to results for the real and imaginary conductivities σ_r and σ_i of

$$\sigma_r = \frac{4}{3} \frac{e^2 \lambda n}{(2\pi m k T)^{\frac{1}{2}}} K_2(x_1) \quad \text{and} \quad \sigma_i = \frac{4}{3} \frac{e^2 \lambda n}{(2\pi m k T)^{\frac{1}{2}}} x_1^{\frac{1}{2}} K_{\frac{3}{2}}(x_1), \quad (8)$$

where λ is the mean free path (deduced from ordinary kinetic theory) and

$$x_1 = \frac{1}{2} \frac{m}{k T} (\omega \lambda)^2$$

$$\text{and} \quad \left. \begin{aligned} K_2(x_1) &= 1 - x_1 - x_1^2 e^{x_1} \text{Ei}(-x_1), \\ K_{\frac{3}{2}}(x_1) &= (\tfrac{1}{2} - x_1) \pi^{\frac{1}{2}} + \pi x_1^{\frac{3}{2}} e^{x_1} (1 - \text{erf}[(x_1)^{\frac{1}{2}}]), \end{aligned} \right\} \quad (9)$$

in which Ei and erf represent the exponential integral and the error function respectively. λ is related to the collision frequency by $\lambda \omega_1 = V$, where V is the mean velocity of the electrons and ω_1 is as defined by (5). The functions (9) have been evaluated using the Jahnke & Emde Funktionentafeln for Ei and erf in order to test this theory.

The results show quite conclusively that there is very little to choose between the simple case and the above from the point of view of dependence of attenuation on n , ω and ω_1 (or λ). Table 1 shows values of σ_r and $\epsilon \left(= 1 - \frac{4\pi\sigma_i}{\omega} \right)$ for the case of 10^{11} electrons per cm.³ at 2000° K with $\lambda = 4 \times 10^{-4}$ cm. (i.e. $\omega_1 = 6.8 \times 10^{10}$).

TABLE 1. COMPARISON OF SIMPLE AND MARGENAU THEORIES OF ATTENUATION

Data: $T = 2000^\circ \text{K}$, $\lambda = 4 \times 10^{-4}$ cm. ($\omega_1 = 6.8 \times 10^{10} \text{sec.}^{-1}$), $n = 10^{11}$ electrons/cm.³.

angular frequency (sec. ⁻¹)	σ_r (Margenau)	σ_r (simple)	ϵ (Margenau)	ϵ (simple)	$\frac{\beta_{\text{Margenau}}}{\beta_{\text{simple}}}$
2×10^{10} (10 cm.)	29.1×10^7	34.1×10^7	0.981	0.947	1.17
6.3×10^{10} (3 cm.)	18.3×10^7	20.8×10^7	0.972	0.968	1.17
16×10^{10} (1.2 cm.)	6.7×10^7	5.7×10^7	0.988	0.988	1.18
25×10^{10} (0.8 cm.)	3.1×10^7	2.6×10^7	0.995	0.994	1.18

The last column gives the ratio of the attenuations predicted on the two theories and is sensibly constant. Similar results are obtained for other values of n and λ . The only significant difference is that on Margenau's theory a given number of electrons will produce an attenuation about 20 % greater than on the simple theory and hence—if the Margenau interpretation is correct—a given attenuation will accord with about 20 % fewer electrons per cm.³. This simple Margenau correction has been adopted in the following sections where electron populations are estimated, since it appears to be the more preferable theory, but since it is not possible from the other data available for the flames to predict n (or $[\epsilon]$, which is n expressed as a partial pressure in atmospheres) to an absolute value which would distinguish between the two theories, this is not very important.

With this section of the theory established it is now possible to examine the behaviour of a large number of alkali metal salts with 3 cm. radiation with confidence

using the value of $\omega_1 = 8.8 \times 10$. Putting in the values of the constants in equation (6) it is possible to write a simple proportionality between n and β , making the 20 % Margenau correction, viz.

$$n = (2.6 \times 10^{11}) \beta. \quad (10)$$

This equation has been used to obtain n at 3 cm. wave-length. Subsequent work will deal with the results obtained with the aid of this equation.

The authors wish to express their gratitude to Professor R. G. W. Norrish for constant advice and encouragement in this work. They are grateful also to the Royal Society for a grant for apparatus and to the Ministry of Supply for the loan of various centimetric wave-length apparatus.

REFERENCES

- Andrew, E., Axford, D. W. E. & Sugden, T. M. 1948 *Trans. Faraday Soc.* **44**, 427.
 Margenau, H. 1945 *Phys. Rev.* **69**, 508.
 Wilson, H. A. 1944 *Modern physics*. London: Blackie, p. 321.

The external gravitational and electromagnetic fields of rotating bodies

BY G. L. CLARK, PH.D.

Department of Mathematics, Bedford College, University of London

(Communicated by Sir Edmund Whittaker, F.R.S.—Received 19 September 1949—
 Revised 16 January 1950)

In previous papers (1941, 1947) the author has shown that the loss of gravitational energy of a rotating system such as a rod may be established by considering the non-periodic terms of order $I^2\omega^5/r^3$ in the equation $G_{\mu\nu} = 0$ for the external field, I being the moment of inertia in the case of a rod.

The present investigation begins by calculating the terms of order $I^2\omega^6/r^2$ in the expressions for $G_{\mu\nu}$. It is found that the equation $G_{\mu\nu} = 0$ has no solution. Values of the $g_{\mu\nu}$ can be determined, however, for which the material energy-tensor has the form of the Faraday-Maxwell stress-tensor due to certain periodic electromagnetic fields. In this way a link is established between periodic gravitational and electromagnetic fields, but the electromagnetic field is not uniquely determined by the analysis.

Next a rotating sphere is considered. The calculation is carried out as far as terms of order $m^3\omega^2/r^7$ in the $G_{\mu\nu}$. Again it is found impossible to find a solution of the equation $G_{\mu\nu} = 0$, but the 'material' energy-tensor can be identified with an electromagnetic energy-tensor.

In both the cases of the periodic system and of the sphere potentials are found for which the energy-tensor represents a field of the form proposed by Blackett (1947), but the investigation does not rule out the possibility that other alternative solutions may exist. The solution in the case of a 'Blackett field' requires a current-charge vector inside the material with a non-zero charge, but the precise values of either the charge or current densities are determined. The solution, in fact, covers the case of a rotating charge as well as of a rotating 'neutral' body. Presumably, further physical considerations must be taken into account to solve the problem completely. As the rotation of the earth produces no visible electric field, the resultant charge must be small.

1. INTRODUCTION

Many of the most interesting and perplexing difficulties in relativity mechanics are concerned with the phenomenon of rotation. The problems may be divided into two groups, namely, those which can be investigated by considering external fields only and those in which the elastic properties of the material must be taken into account. Thus Euler's equations (Clark 1941, § 7) and the loss of gravitational energy of a rotating rod (Clark 1941, § 11) have been derived as conditions of integrability of the equation $G_{\mu\nu} = 0$. On the other hand, in the determination of the free surface of a rotating mass of liquid under isotropic pressure (Clark 1941, § 12) and the radius of a rotating incompressible disk (Clark 1949*a, b*) the internal fields only have to be considered. The second paper on the latter problem successfully dealt with the case of a large angular velocity. A further calculation, however, shows that the theory cannot be extended easily when terms of the second order in the density are retained. The same conclusion applies in the case of a rotating sphere. On the other hand, it appears that a reasonable theory of elasticity can be developed if a rotating body produces an electromagnetic field. The question whether a body can rotate without an associated electromagnetic field, whatever the law of internal constitution, forms the subject-matter of the present paper. External fields only are considered.

A straightforward but tedious calculation shows that the equation $G_{\mu\nu} = 0$ cannot be satisfied. Since the material energy-tensor vanishes outside the body, the result implies that a rotating body produces an electromagnetic field. It is shown that an electromagnetic field of the form proposed by Blackett is possible, at least to the order of approximation considered. The analysis, however, does not rule out the possibility that alternative fields may be found to satisfy the conditions.

In a subsequent paper the extension of the discussion on a rotating mass of liquid under isotropic pressure (Clark 1941, § 12) to a higher order of approximation is described. An attempt to satisfy the equations of motion breaks down. The equations may, however, be satisfied if it is assumed that the rotation produces an electromagnetic field satisfying certain conditions. Once again it is found that a field of the type suggested by Blackett is a possible solution to the problem with a current-charge vector undetermined except for the condition that the charge has a non-zero value.

2. APPROXIMATE VALUES OF THE EINSTEIN TENSOR $G_{\mu\nu}$

As in previous papers we use the convention that Latin indices take only the spatial values 1, 2, 3, while Greek indices refer to both space and time, running over the values 1, 2, 3, 4. We denote the ordinary derivative of a quantity by means of a line followed by the appropriate suffix, as

$$\frac{\partial g_{\mu\nu}}{\partial x_\sigma} \rightarrow g_{\mu\nu|\sigma}, \quad \frac{\partial^2 g_{\mu\nu}}{\partial x_\sigma \partial x_\rho} \rightarrow g_{\mu\nu|\sigma\rho}.$$

Also to avoid unnecessary printing, we adopt units in which both the velocity of light and the constant of gravitation are unity. We set

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu},$$

where the $\eta_{\mu\nu}$ are the Galilean values of $g_{\mu\nu}$. The expressions for $G_{\mu\nu}$ are then

$$G_{\mu\nu} = -\frac{1}{2}M_{\mu\nu} + L_{\mu\nu}, \tag{2.1}$$

where $-\frac{1}{2}M_{\mu\nu}$ represents the linear and $L_{\mu\nu}$ the non-linear terms. We have

$$\left. \begin{aligned} M_{44} &= h_{44|ss} - 2h_{4s|s4} + h_{44|44}, \\ M_{4n} &= h_{4n|ss} - h_{4s|sn} + h_{4n|n4} - h_{ns|s4}, \\ M_{mn} &= h_{mn|ss} - h_{ms|sn} - h_{ns|sm} + h_{4n|mn} - h_{mn|44} + h_{4m|n4} + h_{4n|m4} - h_{44|mn}. \end{aligned} \right\} \quad (2.2)$$

In general the $h_{\mu\nu}$ will be small relative to unity, and if we neglect the cubes of $h_{\mu\nu}$ we can determine the values of $L_{\mu\nu}$ without much difficulty; we find

$$\begin{aligned} L_{44} &= h_{rs}h_{4s|4r} - \frac{1}{2}h_{rs}h_{44|rs} + h_{rs|r}h_{4s|4} - \frac{1}{2}h_{rs|r}h_{44|s} - \frac{1}{2}h_{rs}h_{rs|44} - \frac{1}{2}h_{4r|r}h_{44|4} \\ &\quad - \frac{1}{4}h_{rs|4}h_{rs|4} + \frac{1}{2}h_{4s|r}h_{4r|s} - \frac{1}{2}h_{4r|s}h_{4r|s} + \frac{1}{4}h_{44|s}h_{44|s} - \frac{1}{2}h_{4r|4}h_{ss|r} \\ &\quad + \frac{1}{4}h_{44|r}h_{ss|r} + \frac{1}{4}h_{44|4}h_{ss|4}, \end{aligned} \quad (2.3)$$

$$\begin{aligned} L_{4n} &= \frac{1}{2}h_{rs|r}h_{4s|n} + \frac{1}{2}h_{rs|r}h_{ns|4} - \frac{1}{2}h_{rs|r}h_{4n|s} + \frac{1}{2}h_{rs}h_{4s|nr} + \frac{1}{2}h_{rs}h_{ns|4r} - \frac{1}{2}h_{rs}h_{4n|sr} \\ &\quad - \frac{1}{4}h_{rs|4}h_{rs|n} - \frac{1}{2}h_{rs}h_{rs|n4} - \frac{1}{2}h_{4r|r}h_{44|n} - \frac{1}{2}h_{4r}h_{44|nr} + \frac{1}{2}h_{4r}h_{n4|r4} + \frac{1}{2}h_{4r}h_{4r|n4} \\ &\quad - \frac{1}{2}h_{4r}h_{nr|44} + \frac{1}{2}h_{nr|s}h_{4s|r} - \frac{1}{2}h_{ns|r}h_{4s|4r} - \frac{1}{4}h_{nr|4}h_{44|r} + \frac{1}{4}h_{4n|s}h_{44|s} - \frac{1}{4}h_{4r|n}h_{ss|r} \\ &\quad - \frac{1}{4}h_{nr|4}h_{ss|r} + \frac{1}{4}h_{4n|r}h_{ss|r} + \frac{1}{4}h_{4r|n}h_{44|r} + \frac{1}{4}h_{44|n}h_{rr|4}, \end{aligned} \quad (2.4)$$

$$\begin{aligned} L_{mn} &= \frac{1}{2}h_{rs|r}h_{ms|n} + \frac{1}{2}h_{rs|r}h_{ns|m} - \frac{1}{2}h_{rs|r}h_{mn|s} + \frac{1}{2}h_{rs}h_{ms|nr} + \frac{1}{2}h_{rs}h_{ns|mr} - \frac{1}{2}h_{rs}h_{mn|sr} \\ &\quad - \frac{1}{2}h_{4s|4}h_{ms|n} - \frac{1}{2}h_{4s|4}h_{ns|m} + \frac{1}{2}h_{4s|4}h_{mn|s} - \frac{1}{2}h_{4s}h_{ms|n4} - \frac{1}{2}h_{4s}h_{ns|m4} \\ &\quad + h_{4s}h_{mn|s4} - \frac{1}{2}h_{4r|r}h_{m4|n} - \frac{1}{2}h_{4r|r}h_{n4|m} + \frac{1}{2}h_{4r|r}h_{mn|4} - \frac{1}{2}h_{4r}h_{m4|nr} \\ &\quad - \frac{1}{2}h_{4r}h_{n4|mr} + \frac{1}{4}h_{44|4}h_{m4|n} + \frac{1}{4}h_{44|4}h_{4n|m} - \frac{1}{4}h_{44|4}h_{mn|4} + \frac{1}{2}h_{44}h_{m4|n4} \\ &\quad + \frac{1}{2}h_{44}h_{n4|m4} - \frac{1}{2}h_{44}h_{mn|44} - \frac{1}{4}h_{44|s}h_{mn|s} + \frac{1}{4}h_{44|s}h_{ms|n} + \frac{1}{4}h_{44|s}h_{ns|m} \\ &\quad - \frac{1}{4}h_{rs|n}h_{rs|m} + \frac{1}{2}h_{4s|n}h_{4s|m} - \frac{1}{2}h_{rs}h_{rs|mn} + h_{4s}h_{4s|mn} - \frac{1}{4}h_{44|n}h_{44|m} \\ &\quad - \frac{1}{2}h_{44}h_{44|mn} + \frac{1}{2}h_{mr|s}h_{ns|r} - \frac{1}{2}h_{ms|r}h_{ns|r} - \frac{1}{2}h_{ms|4}h_{n4|s} + \frac{1}{2}h_{ms|4}h_{ns|4} \\ &\quad + \frac{1}{2}h_{4m|s}h_{4n|s} - \frac{1}{2}h_{4m|s}h_{ns|4} - \frac{1}{4}h_{ms|n}h_{rr|s} - \frac{1}{4}h_{ns|m}h_{rr|s} + \frac{1}{4}h_{mn|s}h_{rr|s} \\ &\quad + \frac{1}{4}h_{4m|n}h_{rr|4} + \frac{1}{4}h_{4n|m}h_{rr|4} - \frac{1}{4}h_{mn|4}h_{rr|4}. \end{aligned} \quad (2.5)$$

Instead of the $h_{\mu\nu}$ and $L_{\mu\nu}$, it is often convenient to use quantities $\gamma_{\mu\nu}$, $\Lambda'_{\mu\nu}$ defined by

$$\left. \begin{aligned} \gamma_{\mu\nu} &= h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\eta^{\sigma\rho}h_{\sigma\rho}, \\ \Lambda'_{\mu\nu} &= L_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\eta^{\sigma\rho}L_{\sigma\rho}. \end{aligned} \right\} \quad (2.6)$$

In an important paper Einstein & Infeld (1940) have derived a number of conditions for the existence of a solution of the equation $G_{\mu\nu} = 0$. In particular, they have shown that if $G_{\mu\nu} = 0$, then

$$\int (2\Lambda'_{4n} - \gamma_{44|4n}) \cos(\mathbf{n} \cdot \mathbf{N}) dS = 0, \quad (2.7)$$

where the surface integral is evaluated over any surface S which does not pass through the body and $(\mathbf{n} \cdot \mathbf{N})$ denotes the angle between the direction of x_n and the normal to S and the summation convention applies to the n (Einstein & Infeld 1940, equation (5.16)). In the case of static fields another necessary condition is (Einstein & Infeld 1940, equations (1.10), (3.4))

$$\Lambda'_{mn|n} = 0. \quad (2.8)$$

The condition (2.8) is of fundamental importance in the present investigation.

3. SPECIAL FORMS OF $2\Lambda'_{mn}$ -STATIC FIELDS

Throughout the calculation we require the values of $2\Lambda'_{mn}$ or $2\Lambda'_{mn|n}$. Certain simplified forms taken by these quantities in special cases will now be listed for reference. All potentials are assumed to be independent of the time co-ordinate, and in this section the contributions from the quadratic terms only are considered. In the next section some of the contributions from the cubic terms are given.

(a) If
$$h_{mn} = \phi \delta_{mn} + h'_{mn}, \quad h_{44} = \phi + h'_{44}, \quad h_{4n} = 0,$$

where ϕ is a harmonic function, the expressions (2.3) and (2.5) give for the cross-terms

$$\begin{aligned} 2L_{44} &= -\phi_{|rs} h'_{rs} + \phi_{|s} \left(\frac{3}{2} h'_{44|s} + \frac{1}{2} h'_{u|s} - h'_{rs|r} \right) - \phi h'_{44|ss}, \\ 2L_{mn} &= \phi (h'_{ms|sn} + h'_{ns|sm} - h'_{u|mn} - h'_{mn|ss} - h'_{44|mn}) + \phi_{|s} (h'_{ms|n} + h'_{ns|m} - 2h'_{mn|s}) \\ &\quad + \phi_{|s} \delta_{mn} \left(\frac{1}{2} h'_{u|s} - h'_{rs|r} - \frac{1}{2} h'_{44|s} \right) + \phi_{|n} (h'_{mr|r} - h'_{u|m}) \\ &\quad + \phi_{|m} (h'_{nr|r} - h'_{u|n}) - \phi_{|rs} h'_{rs} \delta_{mn} + \phi_{|ms} h'_{ns} + \phi_{|ns} h'_{ms} - \phi_{|mn} (h'_{u|} + h'_{44}). \end{aligned}$$

When $\phi = -2m/r$, these expressions give

$$\begin{aligned} 2\Lambda'_{mn} &= \frac{2m}{r} (-h'_{ms|sn} - h'_{ns|sm} + h'_{u|mn} + h'_{mn|ss} + h'_{44|mn}) + \frac{2mx_n}{r^3} (h'_{mr|r} - h'_{u|m}) \\ &\quad + \frac{2mx_m}{r^3} (h'_{nr|r} - h'_{u|n}) + \frac{2mx_s}{r^3} (h'_{ms|n} + h'_{ns|m} - 2h'_{mn|s}) + \frac{4m}{r^3} h'_{mn} \\ &\quad - \frac{6mx_m x_s}{r^5} h'_{ns} - \frac{6mx_n x_s}{r^5} h'_{ms} + \frac{6mx_m x_n}{r^5} (h'_{u|} + h'_{44}) \\ &\quad + m\delta_{mn} \left\{ -\frac{1}{r^3} (2h'_{44} + 4h'_{u|}) + 6\frac{x_r x_s}{r^5} h'_{rs} + \frac{2}{r} (h'_{rs|sr} - h'_{u|ss}) \right. \\ &\quad \left. + \frac{x_s}{r^3} (-4h'_{rs|r} + 4h'_{u|s} + 2h'_{44|s}) \right\}, \end{aligned} \quad (3.1)$$

and
$$2\Lambda'_{mn|n} = \frac{2m}{r} h'_{44|mss}. \quad (3.2)$$

It is useful to introduce the quantity $\gamma'_{\mu\nu}$ by means of the equation

$$\gamma'_{\mu\nu} = h'_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \eta^{\sigma\rho} h'_{\sigma\rho}.$$

In the special case when $\gamma'_{mr|r} = 0$, (3.1) may be written in the form

$$\begin{aligned} 2\Lambda'_{mn} &= -2\phi h'_{44|mn} - \phi_{|m} h'_{44|n} - \phi_{|n} h'_{44|m} + \phi_{|s} (\gamma'_{ms|n} + \gamma'_{ns|m} - 2\gamma'_{mn|s}) \\ &\quad + \phi_{|ms} \gamma'_{ns} + \phi_{|ns} \gamma'_{ms} - 2\phi_{|mn} h'_{44} + \delta_{mn} (-\phi_{|rs} \gamma'_{rs} + 3\phi_{|s} h'_{44|s}) \\ &\quad + \phi (-\gamma'_{mn|ss} + \delta_{mn} h'_{44|ss}). \end{aligned} \quad (3.1a)$$

The expression (3.1a) is particularly useful in the present investigation.

The relation (3.2) may be written quite generally in the form

$$2\Lambda'_{mn|n} = -\phi h'_{44|mss}, \quad (3.2a)$$

whatever the value of the harmonic function ϕ .

(b) In the special case when $\gamma_{ms|s} = 0$,

$$L_{44} = -\frac{1}{2}h_{rs}h_{44|rs} + \frac{1}{2}h_{44|s}h_{44|s},$$

$$L_{mn} = \frac{1}{2}h_{rs}(h_{ms|nr} + h_{ns|mr} - h_{mn|sr} - h_{rs|mn}) - \frac{1}{2}h_{44}h_{44|mn} - \frac{1}{4}h_{rs|n}h_{rs|m} \\ - \frac{1}{4}h_{44|n}h_{44|m} + \frac{1}{2}h_{mr|s}h_{ns|r} - \frac{1}{2}h_{ms|r}h_{ns|r}.$$

These give

$$2\Lambda'_{mn} = h_{rs}(h_{ms|nr} + h_{ns|mr} - h_{mn|sr} - h_{rs|mn}) - h_{44}h_{44|mn} - \frac{1}{2}h_{rs|n}h_{rs|m} - \frac{1}{2}h_{44|n}h_{44|m} \\ + h_{mr|s}h_{ns|r} - h_{ms|r}h_{ns|r} + \delta_{mn}(\frac{3}{4}h_{rs|t}h_{rs|t} + \frac{3}{4}h_{44|s}h_{44|s}) \\ - \frac{1}{2}h_{tr|s}h_{ts|r} + \frac{1}{2}h_{rs}h_{rs|tt} + \frac{1}{2}h_{44}h_{44|ss}), \quad (3.3)$$

and

$$2\Lambda'_{mn|n} = h_{rs}(h_{ms|rnn} - \frac{1}{2}h_{rs|mn}) - \frac{1}{2}h_{44}h_{44|mn}. \quad (3.4)$$

The general formula for $2\Lambda'_{mn|n}$ is

$$2\Lambda'_{mn|n} = \gamma_{rs|t}(h_{ms|nt} - h_{rs|rm} - h_{mr|sr} + h_{n|ms} - h_{44|ms}) \\ + h_{rs}(\frac{1}{2}h_{n|msr} - \frac{1}{2}h_{rs|ndt} - \frac{1}{2}h_{44|rsn} - h_{nt|srt} + h_{ms|rtt}) - \frac{1}{2}h_{44}h_{44|ms}. \quad (3.5)$$

This expression is of minor importance in the present investigation; on the other hand, the simple result (3.4) is frequently required.

(c) When $h_{44} = 0$, $h_{mn} = 0$, we have

$$2L_{44} = h_{4s|r}h_{4r|s} - h_{4r|s}h_{4r|s},$$

$$2L_{mn} = -h_{4r}h_{4n|mr} - h_{4r}h_{4m|nr} + h_{4|sn}h_{4s|m} + h_{4m|s}h_{4n|s} + 2h_{4s}h_{4s|mn}.$$

From these expressions we obtain the results

$$2\Lambda'_{mn} = -h_{4r}h_{4m|nr} - h_{4r}h_{4n|mr} + 2h_{4s}h_{4s|mn} + h_{4s|n}h_{4s|m} + h_{4m|s}h_{4n|s} \\ + \delta_{mn}(\frac{1}{2}h_{4s|r}h_{4r|s} - \frac{3}{2}h_{4s|r}h_{4s|r} - h_{4s}h_{4s|rr}), \quad (3.6)$$

and

$$2\Lambda'_{mn|n} = -h_{4r}h_{4m|ssr} + h_{4r}h_{4r|ms}. \quad (3.7)$$

(d) In the above sections various expressions for the quantity $2\Lambda'_{mn}$ have been given. Only one particular form of the quantity $2\Lambda'_{4n} = 2L_{4n}$ is, however, required. When the field is static and

$$h_{44} = \phi, \quad h_{mn} = \phi\delta_{mn},$$

$$2\Lambda'_{4n} = \phi_{|s}h_{4s|n} - h_{4r}\phi_{|nr}. \quad (3.8)$$

In the special case when $\phi = -2m/r$, this reduces to

$$2\Lambda'_{4n} = \frac{2mx_s}{r^3}h_{4s|n} - \frac{2m}{r}h_{4n} + \frac{6mx_n x_r}{r^5}h_{4r}. \quad (3.8a)$$

4. CUBIC CONTRIBUTIONS TO $2\Lambda'_{mn}$

(a) In the case when

$$h_{44} = \phi, \quad h_{mn} = \phi\delta_{mn}, \quad h_{4n} = 0 \quad (\phi \text{ harmonic}),$$

the contribution to $2\Lambda'_{mn}$ from the cubic terms is

$$2\Lambda'_{mn} = 2\phi\phi\phi_{|mn} - \phi\phi_{|m}\phi_{|n} + \frac{3}{2}\phi\phi_{|s}\phi_{|s}\delta_{mn}. \quad (4.1)$$

The expression (4.1) has been taken directly from equation (15.2) of the paper on the gravitational equations by Einstein, Infeld & Hoffmann (1938). Forming the divergence we obtain the result

$$2\Lambda'_{mn|n} = \frac{1}{2}\phi_{|m}\phi_{|s}\phi_{|s} + 6\phi\phi_{|ms}\phi_{|s}. \quad (4.2)$$

In the actual application of (4.2), ϕ is of the form

$$\phi = W + U, \quad \text{where} \quad U = -\frac{2m}{r}$$

and we require the terms of order $m^2 W$. In these circumstances (4.2) gives

$$\begin{aligned} 2\Lambda'_{mn|n} &= \frac{1}{2}W_{|m}U_{|s}U_{|s} + W_{|s}U_{|s}U_{|m} + 6WU_{|ms}U_{|s} + 6W_{|ms}UU_{|s} + 6W_{|s}UU_{|ms} \\ &= m^2 \left\{ -\frac{22}{r^4}W_{|m} + \frac{76x_mx_s}{r^6}W_{|s} - \frac{48x_m}{r^6}W - \frac{24x_s}{r^4}W_{|ms} \right\}. \end{aligned} \quad (4.3)$$

(b) If h_{4n} is not zero as in (a), the contribution of order $\phi h_{4r}h_{4s}$ is

$$\begin{aligned} 2\Lambda'_{mn} &= -\phi_{|nr}h_{4m}h_{4r} - \phi_{|mr}h_{4r}h_{4n} + \phi_{|m}h_{4r}(h_{4r|n} - h_{4n|r}) + \phi_{|n}h_{4r}(h_{4r|m} - h_{4m|r}) \\ &\quad + \phi_{|s}(h_{4m}h_{4n|s} + h_{4n}h_{4m|s}) + \delta_{mn}(\phi_{|sr}h_{4r}h_{4s} - \phi_{|s}h_{4r}h_{4r|s} - \phi h_{4s|r}h_{4s|r} \\ &\quad + \phi h_{4s|r}h_{4r|s}), \end{aligned} \quad (4.4)$$

and

$$2\Lambda'_{mn|n} = 2\phi_{|sr}h_{4r|m}h_{4s} - 2\phi_{|ms}h_{4r}h_{4s|r} + 2\phi h_{4s|rm}h_{4r|s} - 2\phi h_{4s|rm}h_{4s|r}. \quad (4.5)$$

When $\phi = -2m/r$, (4.5) reduces to

$$\begin{aligned} 2\Lambda'_{mn|n} &= m \left(\frac{4}{r^3}h_{4r|m}h_{4r} - \frac{4}{r^3}h_{4m|r}h_{4r} - 12\frac{x_s x_r}{r^5}h_{4s}h_{4r|m} \right. \\ &\quad \left. + 12\frac{x_m x_s}{r^5}h_{4r}h_{4s|r} + \frac{4}{r}h_{4s|rm}h_{4s|r} - \frac{4}{r}h_{4s|rm}h_{4r|s} \right). \end{aligned} \quad (4.6)$$

(c) The last case which we have to consider is when

$$h_{mn} = V\delta_{mn} + h'_{mn}, \quad h_{44} = V, \quad V = -\frac{2m}{r}, \quad h_{4n} = 0.$$

After some calculation we find that the contribution to $2\Lambda'_{mn}$ of order $m^2 h'_{rs}$ is

$$\begin{aligned} 2\Lambda'_{mn} &= VV_{|s}(h'_{ms|n} + h'_{ns|m} - 3h'_{mn|s}) + 2V(V_{|nr}h'_{mr} + V_{|mr}h'_{nr}) - VV_{|n}h'_{u|n} \\ &\quad - VV_{|m}h'_{u|m} + 2V_{|n}V_{|r}h'_{mr} + 2V_{|m}V_{|r}h'_{nr} - 2VV_{|mn}h'_{u|n} - 2V_{|m}V_{|n}h'_{u|n} \\ &\quad - V_{|r}V_{|r}h'_{mn} + \delta_{mn}(-2VV_{|rs}h'_{rs} + \frac{3}{2}V_{|r}V_{|r}h'_{u|n} + 2VV_{|s}h'_{u|s} - V_{|s}V_{|r}h'_{rs}) \\ &= VV_{|s}(\gamma'_{ms|n} + \gamma'_{ns|m} - 3\gamma'_{mn|s}) + 2V(V_{|nr}\gamma'_{mr} + V_{|mr}\gamma'_{nr}) + VV_{|m}\gamma'_{u|n} \\ &\quad + VV_{|n}\gamma'_{u|m} + 2V_{|n}V_{|r}\gamma'_{mr} + 2V_{|m}V_{|r}\gamma'_{nr} - V_{|r}V_{|r}\gamma'_{mn} \\ &\quad + \delta_{mn}(-2VV_{|rs}\gamma'_{rs} - V_{|r}V_{|r}\gamma'_{u|n} - VV_{|s}\gamma'_{u|s} - V_{|s}V_{|r}\gamma'_{rs}) \\ &= m^2 \left\{ -\frac{4x_s}{r^4}(\gamma'_{ms|n} + \gamma'_{ns|m} - 3\gamma'_{mn|s}) - \frac{20}{r^4}\gamma'_{mn} + \frac{32x_n x_r}{r^6}\gamma'_{mr} + \frac{32x_m x_r}{r^6}\gamma'_{nr} \right. \\ &\quad \left. - \frac{4x_m}{r^4}\gamma'_{u|n} - \frac{4x_n}{r^4}\gamma'_{u|m} + \delta_{mn} \left(\frac{4}{r^4}\gamma'_{u|n} - 28\frac{x_r x_s}{r^6}\gamma'_{rs} + 4\frac{x_s}{r^4}\gamma'_{u|s} \right) \right\}. \end{aligned} \quad (4.7)$$

We also have

$$\begin{aligned} 2\Lambda'_{mn|n} &= 2V_{|n}V_{|nr}\gamma'_{mr} - 2V_{|r}V_{|mr}\gamma'_{u|n} - VV_{|ns}\gamma'_{ns|m} + 2V_{|s}V_{|mr}\gamma'_{rs} \\ &= m^2 \left\{ -\frac{8x_r}{r^6}\gamma'_{mr} + \frac{16x_m}{r^6}\gamma'_{u|n} + \frac{4}{r^4}\gamma'_{u|m} - \frac{12x_r x_s}{r^6}\gamma'_{rs|m} - \frac{24x_m x_s x_r}{r^6}\gamma'_{rs} \right\}. \end{aligned} \quad (4.8)$$

5. THE ENERGY-TENSOR

The expressions for $T_{\mu\nu}$ correct to the second order in the density are complicated (see Clark 1941, equation (12.8)). If, however, the terms of the first order in the density satisfy the equation for empty space, then the components of the energy-tensor, correct to the order of the square of the density, are given by the equations

$$\left. \begin{aligned} 16\pi T_{mn} &= \gamma_{mnss} - \gamma_{ms sn} - \gamma_{ns sm} - \gamma_{mn44} + \gamma_{4m4n} + \gamma_{4n4m} \\ &\quad + \delta_{mn}(\gamma_{rsrs} - 2\gamma_{4s4s} + \gamma_{4444}) - 2\Lambda'_{mn}, \\ 16\pi T_{4n} &= \gamma_{4nss} - \gamma_{ns4s} + \gamma_{444n} - \gamma_{4s4n} - 2\Lambda'_{4n}, \\ 16\pi T_{44} &= \gamma_{44ss} - \gamma_{rsrs} - 2\Lambda'_{44}. \end{aligned} \right\} \quad (5.1)$$

If we impose co-ordinate conditions

$$\gamma_{4\mu4} - \gamma_{\mu s s} = 0, \quad (5.2)$$

the linear terms in the expansion for the energy-tensor are

$$16\pi T_{\mu\nu} = \gamma_{\mu\nu ss} - \gamma_{\mu\nu 44}. \quad (5.3)$$

These equations are sufficient to discuss periodic fields.

In the investigation on a sphere, we consider terms involving the cube of the density; also the components of the energy-tensor of the order of the square of the density are not zero, but are the classical values of the Faraday-Maxwell stress-tensor of certain electromagnetic fields. An amended form of (5.1) will now be derived to cope with this situation in the case of static fields.

We have the fundamental equations

$$\begin{aligned} -8\pi T_{mn} &= G_{mn} - \frac{1}{2}g_{mn}G, \\ -8\pi T_{44} &= G_{44} - \frac{1}{2}g_{44}G. \end{aligned}$$

If $G = 0$, these equations may be written in the form

$$\begin{aligned} -8\pi T_{mn} &= G_{mn} + \frac{1}{2}G\delta_{mn}, \\ -8\pi T_{44} &= G_{44} - \frac{1}{2}G. \end{aligned}$$

$$\begin{aligned} \text{Also, we have } -8\pi T &= G = g^{rs}G_{rs} + 2g^{4s}G_{4s} + g^{44}G_{44} \\ &= -G_{44} + G_{44} - 8\pi(h^{rs}E_{rs}^{(1)} + 2h^{4s}E_{4s}^{(1)} + h^{44}E_{44}^{(1)}), \end{aligned}$$

where $E_{\mu\nu}^{(1)}$ represents the leading values of the components of the energy-tensor. The condition $G = 0$ implies the equation

$$E_{44}^{(1)} - E_{44}^{(1)} = 0.$$

Writing

$$h_{44} = \phi, \quad h_{rs} = \phi\delta_{rs} + h'_{rs}$$

we get

$$-8\pi T = -G_{44} + G_{44} + 16\pi\phi E_{44}^{(1)} + 8\pi h'_{rs} E_{rs}^{(1)} - 16\pi h_{4s} E_{4s}^{(1)}.$$

Accordingly, the expression for T_{mn} assumes the form

$$\begin{aligned} 16\pi T_{mn} &= -(2G_{mn} - G_{44}\delta_{mn} + G_{44}\delta_{mn}) - 16\pi\phi E_{44}^{(1)} - 8\pi h'_{rs} E_{rs}^{(1)} + 16\pi h_{4s} E_{4s}^{(1)} \\ &= M^* - 2\Lambda'_{mn} + 16\pi\delta_{mn}(h_{4s} E_{4s}^{(1)} - \phi E_{44}^{(1)} - \frac{1}{2}h'_{rs} E_{rs}^{(1)}), \end{aligned} \quad (5.4)$$

where

$$M_{mn}^* = \gamma_{mn|es} - \gamma_{ms|en} - \gamma_{ns|em} + \gamma_{rs|er} \delta_{mn}.$$

Since $M_{mn|n}^* = 0$, we have from (5.4)

$$16\pi T_{mn\ n} = -2\Lambda'_{mn\ n} + 16\pi(h_{4s}E_{4s}^{(1)} - \phi E_{44}^{(1)} - \frac{1}{2}h'_{rs}E_{rs}^{(1)})_m. \quad (5.5)$$

When $E_{\mu\nu}^{(1)}$ is zero, the expression (5.4) reduces to the first member of (5.1) when the potentials are independent of the time. In our application of (5.4) the linear terms M_{mn}^* and the quadratic and cubic terms $2\Lambda'_{mn}$ are of the order of the cube of the density, $E_{\mu\nu}^{(1)}$ is of the order of the square of the density and ϕ , h'_{rs} , h_{4s} are of the first order in the density.

6. FIELD DUE TO ROTATING NEARLY SPHERICAL BODY

It has been shown (Clark 1947, § 7) that at great distances the field due to a rotating rod has the same form as the field due to a rotating nearly spherical body. To the order of approximation required the field in terms of the $\gamma_{\mu\nu}$ is given by

$$\left. \begin{aligned} \gamma_{44} &= -\frac{4m}{r} + Ip^2 \left[\frac{(x_1^2 - x_2^2)}{r^3} \cos p(t-r) + \frac{2x_1x_2}{r^3} \sin p(t-r) \right], \\ \gamma_{41} &= -\frac{2A\omega x_2}{r^3} + Ip^2 \left[-\frac{x_1}{r^2} \cos p(t-r) - \frac{x_2}{r^2} \sin p(t-r) \right], \\ \gamma_{42} &= \frac{2A\omega x_1}{r^3} + Ip^2 \left[-\frac{x_1}{r^2} \sin p(t-r) + \frac{x_2}{r^2} \cos p(t-r) \right], \\ \gamma_{11} &= \frac{Ip^2}{r} \cos p(t-r), \\ \gamma_{22} &= -\frac{Ip^2}{r} \cos p(t-r), \\ \gamma_{12} &= \frac{Ip^2}{r} \sin p(t-r), \end{aligned} \right\} \quad (6.1)$$

where in the case of a nearly spherical body of constant density ρ , the boundary is given by

$$r = a \left[1 - \frac{\epsilon}{4a^2} \{ (x_1^2 + x_2^2) + (x_1^2 - x_2^2) \cos pt + 2x_1x_2 \sin pt \} \right], \quad (6.2)$$

and

$$m = \frac{4}{3}\pi\rho a^3(1 - \frac{1}{2}\epsilon), \quad A = \frac{8\pi\rho a^5}{15}(1 - \epsilon),$$

$$I = -\frac{4\pi\rho\epsilon a^5}{15}, \quad p = 2\omega.$$

In the case of a rod $A = I$.

It is important to remember, moreover, that the field (6.1) also applies to the orbital motion of two particles (Clark 1946, § 3) with $A = 2I$, $I = mb^2$ in the case of particles of equal mass m moving in a circle of radius b about their centre of gravity. The results of this section accordingly apply equally well to orbital motion.

The phenomenon of the loss of gravitational energy has been established by substituting the periodic potentials in the expressions (2.3) and (2.5) and picking out the non-periodic terms of order $I^2\omega^5/r^3$. We now carry out the corresponding analysis for the terms of order $I^2\omega^6/r^2$.

As in the previous calculation we obtain terms with factors

$$\sin p(t-r) \cos p(t-r), \quad \cos^2 p(t-r) \quad \text{and} \quad \sin^2 p(t-r).$$

We may now write

$$\begin{aligned} \cos p(t-r) \sin p(t-r) &= \frac{1}{2} \sin 2p(t-r), \\ \cos^2 p(t-r) &= \frac{1}{2}(1 + \cos 2p(t-r)), \\ \sin^2 p(t-r) &= \frac{1}{2}(1 - \cos 2p(t-r)), \end{aligned}$$

and in this way split up $2L_{\mu\nu}$ into periodic and non-periodic parts. After some calculation the four-periodic terms of order $I^2 p^6/r^2$ are found to be

$$\begin{aligned} 2L_{44} &= \frac{I^2 p^6}{8} \left(\frac{1}{r^2} + \frac{6x_3^2}{r^4} + \frac{x_3^4}{r^6} \right), \\ 2L_{mn} &= \frac{I^2 p^6 x_m x_n}{8} \left(\frac{1}{r^4} + \frac{6x_3^2}{r^6} + \frac{x_3^4}{r^8} \right), \\ 2L_{4n} \frac{x_n}{r} &= -\frac{I^2 p^6}{8} \left(\frac{1}{r^2} + \frac{6x_3^2}{r^4} + \frac{x_3^4}{r^6} \right). \end{aligned}$$

Since $L_{44} = L_{\bar{u}}$, we have $2L_{44} = 2\Lambda'_{44}$, $2L_{mn} = 2\Lambda'_{mn}$.

We attempt to build up a solution of the equation $G_{\mu\nu}$ by taking potentials of the form

$$\begin{aligned} \gamma_{mn} &= \frac{I^2 p^6}{8} \left\{ x_m x_n \left(\frac{A_1}{r^2} + \frac{A_2 x_3^2}{r^4} + \frac{A_3 x_3^4}{r^6} \right) + \frac{D_1 x_3^2}{r^2} \delta_{mn} + \frac{D_2 x_3^2}{r^2} \delta_{m3} \delta_{n3} + \frac{D_3 x_3^4}{r^4} \delta_{mn} \right. \\ &\quad \left. + \frac{D_4 x_3}{r^2} (x_n \delta_{m3} + x_m \delta_{n3}) + \frac{D_5 x_3^3}{r^4} (x_m \delta_{n3} + x_n \delta_{m3}) \right\}, \\ \gamma_{44} &= \frac{I^2 p^6}{8} \left(\frac{B_1 x_3^2}{r^2} + \frac{B_2 x_3^4}{r^4} \right), \end{aligned} \quad (6.3)$$

where the A_i , D_i and B_i are disposable constants.

On working out the value of M_{mn}^* it is found that the coefficients D_2 , D_3 and D_5 must vanish and that $D_4 = -2D_1$. The values of the components of the energy-tensor (5.1) are then given by

$$\begin{aligned} 16\pi T_{44} &= \frac{I^2 p^6}{8} \left\{ (2B_1 - 2A_1 + 6D_1 - 1) \frac{1}{r^2} \right. \\ &\quad \left. + (12B_2 - 6B_1 - 2A_2 - 10D_1 - 6) \frac{x_3^2}{r^4} - (20B_2 - 2A_3 - 1) \frac{x_3^4}{r^6} \right\}, \end{aligned} \quad (6.4)$$

$$\begin{aligned} 16\pi T_{mn} &= \frac{I^2 p^6 x_m x_n}{8} \left\{ (2A_1 + 2A_2 - 8D_1 - 1) \frac{1}{r^4} \right. \\ &\quad \left. + (16D_1 - 4A_2 + 12A_3 - 6) \frac{x_3^2}{r^6} - (18A_3 + 1) \frac{x_3^4}{r^8} \right\}. \end{aligned} \quad (6.5)$$

The coefficient D_1 is quite superfluous, for on putting

$$A_2 = 4D_1 + A'_2, \quad B_1 = -3D_1 + B'_1,$$

the relations (6.4) and (6.5) reduce to

$$16\pi T_{44} = \frac{I^2 p^6}{8} \left\{ (2B'_1 - 2A_1 - 1) \frac{1}{r^2} + (12B_2 - 6B'_1 - 2A'_2 - 6) \frac{x_3^2}{r^4} - (20B_2 + 2A_3 + 1) \frac{x_3^4}{r^6} \right\}, \quad (6.6)$$

$$16\pi T_{mn} = \frac{I^2 p^6 x_m x_n}{8} \left\{ (2A_1 + 2A'_2 - 1) \frac{1}{r^4} + (12A_3 - 4A'_2 - 6) \frac{x_3^2}{r^6} - (18A_3 + 1) \frac{x_3^4}{r^8} \right\}. \quad (6.7)$$

It is not possible to choose the constants in (6.6) and (6.7) so that T_{44} and T_{mn} vanish completely. The constants can be chosen, however, so that the energy-tensor has the form of the Faraday-Maxwell stress-energy tensor due to certain electromagnetic fields. For example, if

$$A_3 = -\frac{5}{18}, \quad B'_1 = -\frac{2}{3}, \quad A_1 = \frac{5}{6}, \quad A'_2 = -\frac{7}{3}, \quad B_2 = -\frac{2}{9},$$

then

$$\left. \begin{aligned} 16\pi T_{44} &= -\frac{1}{2} I^2 p^6 \left(\frac{1}{r^2} - \frac{x_3^4}{r^6} \right), \\ 16\pi T_{mn} &= -\frac{1}{2} I^2 p^6 x_m x_n \left(\frac{1}{r^4} - \frac{x_3^4}{r^8} \right). \end{aligned} \right\} \quad (6.8)$$

In § 11 we show that (6.8) represents the energy-tensor due to a constant change of density $\sigma = 2\rho$.

It should be noted that the failure to find a solution of $G_{\mu\nu} = 0$ arises from the fact that the equation

$$\nabla^2 V = 1/r^2$$

has no solution which is finite as $r \rightarrow \infty$. The expressions (6.3) do, in point of fact, satisfy the condition (2.8).

The result has been derived by starting with the *retarded* potentials (6.1). The analysis could be repeated if we started with potentials similar in form to (6.1) but with $\cos p(t-r)$ and $\sin p(t-r)$ replaced respectively by

$$\frac{1}{2} \{ \cos p(t-r) + \cos p(t+r) \} \quad \text{and} \quad \frac{1}{2} \{ \sin p(t-r) + \sin p(t+r) \}.$$

I have not carried out this calculation, as it would merely give the same result.

7. THE MOMENTUM TERMS

As far as the main development of the argument is concerned it is not necessary to consider the momentum terms. They are, nevertheless, rather interesting and well worth investigation. After some calculation the non-periodic terms of order $I^2 p^6 / r^2$ in the expression for $2\Lambda'_{4n} x_n / r$ are found to be given by the relation

$$2\Lambda'_{4n} \frac{x_n}{r} = -\frac{I^2 p^6}{8} \left(\frac{1}{r^2} + \frac{6x_3^2}{r^4} + \frac{x_3^4}{r^6} \right).$$

Therefore

$$\int 2\Lambda'_{4n} \frac{x_n}{r} dS \neq 0,$$

and accordingly it is not possible to find values of γ_{4n} for which the momentum components T_{4n} vanish. We could, however, consider distributions of matter for

which the quantity m in (6.1) is replaced by $m(1 + \alpha a^2 \omega^2)$, where α is a constant. For these distributions of matter we obtain

$$\gamma_{44|4n} = -2m\alpha\omega\dot{\omega}x_n/r^3$$

and
$$\gamma_{44|4n}x_n/r = -2m\alpha\omega\dot{\omega}/r^2 = 32I^2\omega^6\alpha/r^2 = \alpha I^2p^6/2r^2,$$

on using Clark 1947, equation (7.19). By a suitable choice of density distribution we would obtain a value of α which would ensure the vanishing of the surface integral. There are therefore two kinds of systems; one satisfies the condition (2.7), the other does not. The physical interpretation of these two types will be given in § 12.

8. FIELD DUE TO ROTATING SPHERE

The leading terms in the expressions for the internal and external fields of rotating spheres of various density distributions have been given in previous papers. For the sake of definiteness the derivatives of all the $\gamma_{\mu\nu}$ were taken to be continuous at the boundary. This restriction is not at all necessary, the derivative of h_{44} must certainly be continuous, but it does not matter whether or not the $h_{mn|s}$ are continuous. In this paper we take the external field in the general form

$$\gamma_{mn} = \gamma_{mn}^{(1)} + \gamma_{mn}^{(2)} + \gamma_{mn}^{(3)},$$
$$\gamma_{mn}^{(1)} = B_1 \left\{ -\frac{2}{r^3} \delta_{mn} + \frac{2}{r^3} \delta_{m3} \delta_{n3} + \frac{3x_m x_n}{r^5} + \frac{3x_3^2}{r^5} \delta_{mn} - \frac{3x_3}{r^5} (x_n \delta_{m3} + x_m \delta_{n3}) \right\},$$
$$\gamma_{mn}^{(2)} = B_2 \left\{ \frac{1}{3r^3} \delta_{mn} - \frac{2}{3r^3} \delta_{m3} \delta_{n3} - \frac{x_3^2}{r^5} \delta_{mn} + \frac{x_3}{r^5} (x_n \delta_{m3} + x_m \delta_{n3}) \right\},$$
$$\gamma_{mn}^{(3)} = B_3 \left\{ \frac{x_3^2}{r^5} \delta_{mn} + \frac{x_3}{r^5} (x_n \delta_{m3} + x_m \delta_{n3}) - \frac{5x_3^2 x_m x_n}{r^7} \right\},$$
$$\gamma_{4n}^{(1)} = \frac{A}{r^3} (-x_2 \delta_{n1} + x_1 \delta_{n2}),$$
$$\gamma_{44} = -\frac{4m}{r} - \gamma_{ll} + 2B_4 \left(\frac{1}{3r^3} - \frac{x_3^2}{r^5} \right),$$

}

(8.1)

where $A = A' ma^2 \omega$, $B_i = B'_i ma^4 \omega^2$ and A' , B'_i are numerical factors.

From the above we have

$$h_{44} = -\frac{2m}{r} + B_4 \left(\frac{1}{3r^3} - \frac{x_3^2}{r^5} \right).$$

Since both h_{44} and $h_{44|s}$ are continuous at the boundary, it follows that if the pressure and density are functions of r only, then B_4 cannot be zero. Nor can it be zero in the case of a rotating mass of liquid where the form of the free surface is a spheroid. In the first instance we may put B_4 equal to zero and discuss the remaining terms. A lengthy and tedious calculation reveals that with B_4 zero, the equation $G_{\mu\nu} = 0$ can be solved for arbitrary values of A' , B'_i ($i = 1, 2, 3$). To avoid unnecessary printing this part of the investigation will be omitted and terms involving B_4 only will be considered.

9. THE TERMS IN B_4

The terms of order m^2 and independent of ω have occurred in the investigation on the motion of n bodies. We have (Clark 1941, pp. 242-243),

$$\gamma'_{44} = -\frac{3m^2}{r^2}, \quad \gamma'_{mn} = \frac{7m^2 x_m x_n}{r^4}, \quad \gamma'_{ms|s} = 0.$$

In terms of the $h_{\mu\nu}$, these values give

$$h'_{44} = \frac{2m^2}{r^2}, \quad h'_{mn} = \frac{7m^2 x_m x_n}{r^4} - \frac{5m^2}{r^2} \delta_{mn}. \quad (9.1)$$

We consider potentials of the form

$$h^{(1)}_{44} = \phi, \quad h^{(1)}_{mn} = \phi \delta_{mn}, \quad (9.2)$$

$$\text{where} \quad \phi = -\frac{2m}{r} + W, \quad W = B_4 \left(\frac{1}{3r^3} - \frac{x_3^2}{r^5} \right), \quad B_4 = B'_4 m a^4 \omega^2 \quad (9.3)$$

and B'_4 is a numerical constant.

The values of $2L_{44}$, $2\Lambda'_{mn}$ of order $m^2\omega^2/r^6$ are

$$2L_{44} = mB_4 \left(-\frac{4}{r^6} + \frac{12x_3^2}{r^8} \right),$$

$$2\Lambda'_m = mB_4 \left\{ -\frac{34}{3r^6} \delta_{mn} + \frac{28x_m x_n}{r^8} + \frac{44x_3}{r^8} (x_m \delta_{n3} + x_n \delta_{m3}) \right. \\ \left. + \frac{42x_3^2}{r^8} \delta_{mn} - \frac{8}{r^6} \delta_{m3} \delta_{n3} - \frac{172x_3^2 x_m x_n}{r^{10}} \right\}.$$

The equations

$$h''_{44|ss} = 2L_{44}, \quad M^*_{mn} = 2\Lambda'_{mn},$$

have solutions

$$h''_{44} = mB_4 \left(-\frac{2}{3r^4} + \frac{2x_3^2}{r^6} \right),$$

$$\gamma''_{mn} = mB_4 \left\{ -\frac{1}{2r^4} \delta_{mn} - \frac{5x_m x_n}{2r^6} - \frac{7x_3}{r^6} (x_m \delta_{n3} + x_n \delta_{m3}) \right. \\ \left. - \frac{x_3^2}{6r^6} \delta_{mn} + \frac{5}{3r^4} \delta_{m3} \delta_{n3} + \frac{43x_3^2 x_m x_n}{2r^8} \right\}.$$

The contributions to $2\Lambda'_{mn}$ of order $m^3\omega^2/r^7$ arise from three sources; they are

- (i) $2\Lambda^{(1)}_{mn}$ on substituting h''_{44} , γ''_{mn} in (3.1),
- (ii) $2\Lambda^{(2)}_{mn}$ on substituting $h^{(1)}_{44} + h'_{44}$, $h^{(1)}_{mn} + h'_{mn}$ in (3.3), where h'_{44} , h'_{mn} are given by (9.1),
- (iii) $2\Lambda^{(3)}_{mn}$ on using (4.1).

For present purposes it is sufficient to calculate the divergence $2\Lambda'_{mn|n}$.

From (3.2a) we have

$$2\Lambda^{(1)}_{mn|n} = \frac{2m}{r} h''_{44|mss}$$

$$= m^2 B_4 \left(\frac{48x_m}{r^9} + \frac{48x_3}{r^9} \delta_{m3} - \frac{192x_3^2 x_m}{r^{11}} \right),$$

$$2\Lambda^{(2)}_{mn|n} = -W h'_{44|mss}$$

$$= m^2 B_4 \left(\frac{16x_m}{3r^9} - \frac{16x_m x_3^2}{r^{11}} \right).$$

Finally, from (4.3) we have

$$2\Lambda_{mn|n}^{(3)} = m^2 B_4 \left(-\frac{166x_m}{r^9} - \frac{148x_3}{r^9} \delta_{m3} + \frac{646x_3^2 x_m}{r^{11}} \right).$$

Adding these expressions, we get

$$2\Lambda_{mn|n} = -m^3 a^4 \omega^2 \left(\frac{P'x_m}{r^9} + \frac{Q'x_3}{r^9} \delta_{m3} + \frac{R'x_m x_3^2}{r^{11}} \right), \quad (9.4)$$

where

$$P' \cong \frac{338}{3} B'_4, \quad Q' = 100 B'_4, \quad R' = -438 B'_4. \quad (9.5)$$

The non-vanishing of the expression on the right of (9.4) is of fundamental importance. First, it shows that the equation $G_{\mu\nu} = 0$ has no solution and that we must associate an electromagnetic field with the rotation. Secondly, it requires that the current-charge vector should not be zero outside the sphere when terms of order $m^3 \omega^3$ are considered. It will be remembered that a similar phenomenon occurs in Weyl's theory (Eddington 1924, p. 211).

In the next half of the paper we discuss the field due to various current-charge distributions. The first part of the investigation is at the classical level of approximation.

10. MAXWELL'S EQUATIONS

In the classical theory the electromagnetic field is described by vectors $\mathbf{E}$ and $\mathbf{H}$. The vector $\mathbf{E}$ with components E_i represents the electric field and the vector $\mathbf{H}$ (H_i) represents the magnetic field. Let ρ represent the density of electric charge and σ (σ_i) the density of electric current.

The laws of electromagnetism are those given by Maxwell. These are

$$\left. \begin{aligned} \text{curl } \mathbf{E} &= -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, & \nabla \cdot \mathbf{E} &= 4\pi\rho, \\ \text{curl } \mathbf{H} &= \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \boldsymbol{\sigma}, & \nabla \cdot \mathbf{H} &= 0. \end{aligned} \right\} \quad (10.1)$$

In the above equations the electrostatic unit of charge has been used. In order to avoid unnecessary printing we shall take $c = 1$ as in the first half of the paper. By using the Heaviside-Lorentz unit of charge we could avoid the factor 4π , but we shall not adopt these units.

If the four-dimensional current-charge vector $J^\mu = (\sigma_1, \sigma_2, \sigma_3, \rho)$ is zero, both $\mathbf{E}$ and $\mathbf{H}$ satisfy the equations

$$\nabla^2 \mathbf{A} - \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0, \quad \nabla \cdot \mathbf{A} = 0. \quad (10.2)$$

In the next section we shall calculate the values of $\mathbf{H}$ and $\mathbf{E}$ for a distribution of constant charge in the form of a nearly spherical body rotating about an axis other than that of symmetry. We then calculate the covariant components of the Faraday-Maxwell stress-tensor. These are given by

$$\left. \begin{aligned} 16\pi E_{mn} &= 2[\delta_{mn}(\mathbf{E}^2 + \mathbf{H}^2) - 2H_m H_n - 2E_m E_n], \\ 16\pi E_{4n} &= 4\epsilon_{nik} H_i E_k, \\ 16\pi E_{44} &= 2(\mathbf{E}^2 + \mathbf{H}^2), \end{aligned} \right\} \quad (10.3)$$

where ϵ_{nik} is the Levi-Civita tensor density. In §§ 13, 14 we describe the relativistic generalization of the classical theory.

11. FIELD DUE TO NEARLY SPHERICAL CHARGE DISTRIBUTION

When the fields are changing very slowly the equations (10.1) and (10.2) reduce respectively to

$$\left. \begin{aligned} \text{curl } \mathbf{E} &= 0, & \nabla \mathbf{E} &= 4\pi\rho, \\ \text{curl } \mathbf{H} &= 4\pi, & \nabla \mathbf{H} &= 0, \end{aligned} \right\} \quad (11.1)$$

and
$$\nabla^2 \mathbf{A} = 0, \quad \nabla \mathbf{A} = 0. \quad (11.2)$$

We first take the values of $\mathbf{E}$ and $\mathbf{H}$ for a sphere of constant charge which will be denoted by σ . We then add further terms so that $\mathbf{E}$ and $\mathbf{H}$ are continuous at the boundary (6.2). In this way we obtain all the external fields

$$\begin{aligned} H_1 &= \frac{4\pi\sigma a^5 \omega x_1 x_3}{5r^5} + \pi\sigma\epsilon\omega \left\{ \sin pt \left(-\frac{2x_2 x_3 a^5}{5r^5} + \frac{2x_2 x_3 a^7}{7r^7} - \frac{2x_1^2 x_2 x_3 a^7}{r^9} \right) \right. \\ &\quad \left. + \cos pt \left(-\frac{2x_1 x_3 a^5}{5r^5} + \frac{2x_1 x_3 a^7}{7r^7} - \frac{x_1 x_3 (x_1^2 - x_2^2) a^7}{r^9} \right) \right\}, \\ H_2 &= \frac{4\pi\sigma a^5 \omega x_2 x_3}{5r^5} + \pi\sigma\epsilon\omega \left\{ \sin pt \left(-\frac{2x_1 x_3 a^5}{5r^5} + \frac{2x_1 x_3 a^7}{7r^7} - \frac{2x_1 x_2^2 x_3 a^7}{r^9} \right) \right. \\ &\quad \left. + \cos pt \left(\frac{2x_2 x_3 a^5}{5r^5} - \frac{2x_2 x_3 a^7}{7r^7} - \frac{x_2 x_3 (x_1^2 - x_2^2) a^7}{r^9} \right) \right\}, \\ H_3 &= \frac{4\pi\sigma\omega a^5}{5} \left(-\frac{1}{3r^3} + \frac{x_3^2}{r^5} \right) \\ &\quad + \pi\sigma\omega\epsilon \left\{ \sin pt \left(\frac{4x_1 x_2 a^5}{5r^5} + \frac{2x_1 x_2 a^7}{7r^7} - \frac{2x_1 x_2 x_3^2 a^7}{r^9} \right) \right. \\ &\quad \left. + \cos pt \left(\frac{2(x_1^2 - x_2^2) a^5}{5r^5} + \frac{(x_1^2 - x_2^2) a^7}{7r^7} - \frac{x_3^2 (x_1^2 - x_2^2) a^7}{r^9} \right) \right\}, \\ E_1 &= \frac{4\pi\sigma a^3 x_1}{3r^3} + \pi\sigma\epsilon a^5 \left\{ \sin pt \left(\frac{2x_2}{5r^5} - \frac{2x_1^2 x_2}{r^7} \right) + \cos pt \left(\frac{2x_1}{5r^5} - \frac{x_1 (x_1^2 - x_2^2)}{r^7} \right) \right\}, \\ E_2 &= \frac{4\pi\sigma a^3 x_2}{3r^3} + \pi\sigma\epsilon a^5 \left\{ \sin pt \left(\frac{2x_1}{5r^5} - \frac{2x_1 x_2^2}{r^7} \right) + \cos pt \left(-\frac{2x_2}{5r^5} - \frac{x_2 (x_1^2 - x_2^2)}{r^7} \right) \right\}, \\ E_3 &= \frac{4\pi\sigma a^3 x_3}{3r^3} - \pi\sigma\epsilon a^5 \left\{ \frac{2x_1 x_2 x_3}{r^7} \sin pt + \frac{x_3 (x_1^2 - x_2^2)}{r^7} \cos pt \right\}. \end{aligned}$$

In the above expressions terms of order $\sigma\omega^2$ have been neglected. Different fields would be produced by charges of the form

$$\sigma + A\epsilon \cos pt + B\epsilon \sin pt. \quad (11.3)$$

We have only taken the charge to be constant for the sake of simplicity and for no other reason. The next step is to determine, by successive approximation, the terms

of higher order in ω . As far as the present investigation is concerned we only require the leading terms at great distances. These are automatically fixed by the leading terms in the expressions given above. The calculation is not at all difficult as the general forms of $\mathbf{E}$ and $\mathbf{H}$ can be guessed easily enough. Retaining terms up to order $\sigma\omega^6$ we find a solution of (10.1) for which the current charge vector is zero,

$$H_1 = -\chi(x_2x_3\sin pt + x_1x_3\cos pt),$$

$$H_2 = \chi(-x_1x_3\sin pt + x_2x_3\cos pt),$$

$$H_3 = \chi(2x_1x_2\sin pt + (x_1^2 - x_2^2)\cos pt),$$

where

$$\chi = \frac{\pi\sigma\epsilon a^5}{5} \left(\frac{p}{r^5} + \frac{p^3}{6r^3} + \frac{p^5}{24r} \right),$$

and

$$\begin{aligned} E_1 = \pi\sigma\epsilon a^5 \bigg[\sin pt \bigg\{ \frac{2x_2}{5r^5} - \frac{2x_1^2x_2}{r^7} - \frac{x_1^2x_2}{5r^5} p^2 - \frac{1}{60} \left(\frac{x_2}{r} + \frac{x_1^2x_2}{r^3} \right) p^4 + \frac{1}{360} \left(2x_2r - \frac{x_1^2x_2}{r} \right) p^6 \bigg\} \\ + \cos pt \bigg\{ \frac{2x_1}{5r^5} - \frac{x_1(x_1^2 - x_2^2)}{r^7} - \frac{x_1(x_1^2 - x_2^2)}{10r^5} p^2 - \frac{1}{120} \left(\frac{2x_1}{r} + \frac{x_1(x_1^2 - x_2^2)}{r^3} \right) p^4 \\ + \frac{1}{720} \left(4x_1r - \frac{x_1(x_1^2 - x_2^2)}{r} \right) p^6 \bigg\} \bigg], \end{aligned}$$

$$\begin{aligned} E_2 = \pi\sigma\epsilon a^5 \bigg[\sin pt \bigg\{ \frac{2x_1}{5r^5} - \frac{2x_1x_2^2}{r^7} - \frac{x_1x_2^2}{5r^5} p^2 - \frac{1}{60} \left(\frac{x_1}{r} + \frac{x_1x_2^2}{r^3} \right) p^4 + \frac{1}{360} \left(2x_1r - \frac{x_1x_2^2}{r} \right) p^6 \bigg\} \\ + \cos pt \bigg\{ -\frac{2x_2}{5r^5} - \frac{x_2(x_1^2 - x_2^2)}{r^7} - \frac{x_2(x_1^2 - x_2^2)}{10r^5} p^2 + \frac{1}{120} \left(\frac{2x_2}{r} - \frac{x_2(x_1^2 - x_2^2)}{r^3} \right) p^4 \\ + \frac{1}{720} \left(-4x_2r - \frac{x_2(x_1^2 - x_2^2)}{r} \right) p^6 \bigg\} \bigg], \end{aligned}$$

$$E_3 = \pi\sigma\epsilon a^5 [-2x_1x_2x_3\sin pt - x_3(x_1^2 - x_2^2)\cos pt] \mu,$$

where

$$\mu = \frac{1}{r^7} + \frac{1}{10r^5} p^2 + \frac{1}{120r^3} p^4 + \frac{1}{720r} p^6.$$

In the above the fields have been expressed in terms of the 'present' values of the periodic terms; when considering external fields, it is often convenient to use 'retarded' values. Indeed, in the first part of the paper we have already used 'retarded' instead of 'present' values. We can take either the 'present' or 'retarded' solution and substitute in (10.3). Proceeding as in § 6 we can express $E_{\mu\nu}$ in two parts, one non-periodic and the other periodic. As far as the components E_{44} , E_{mn} are concerned the value of the non-periodic part is the same whatever solution is used. We shall see in § 12 that the non-periodic part of the momentum terms E_{4n} may, or may not, be the same. In the case of the above field there is no contribution to E_{4n} of order I^2p^6/r^2 whatever, if present values are used; if retarded values are used,

there are contributions but these cancel out. The components of $E_{\mu\nu}$ of the order considered are found to have the values

$$\begin{aligned} 16\pi E_{mn} &= \frac{2\pi^2}{225} \sigma^2 \epsilon^2 a^{10} p^6 x_m x_n \left(\frac{1}{r^4} - \frac{x_3^4}{r^8} \right) \\ &= \frac{1}{8} \left(\frac{\sigma}{\rho} \right)^2 I^2 p^6 x_m x_n \left(\frac{1}{r^4} - \frac{x_3^4}{r^8} \right), \\ 16\pi E_{44} &= \frac{2\pi^2}{225} \sigma^2 \epsilon^2 a^{10} p^6 \left(\frac{1}{r^2} - \frac{x_3^4}{r^6} \right) \\ &= \frac{1}{8} \left(\frac{\sigma}{\rho} \right)^2 I^2 p^6 \left(\frac{1}{r^2} - \frac{x_3^4}{r^6} \right), \\ 16\pi E_{4n} &= 0, \end{aligned}$$

where, as in § 6,

$$I = -4\pi\rho\epsilon a^5/15.$$

We have thus (except for the reversal of sign) identified the energy-tensor (6.8) with an electromagnetic energy-tensor representing the field due to a rotating distribution of charge. By giving different values to the disposable constants in (6.6) and (6.7) the energy-tensor may be made to represent the field due to a current-charge vector of the form $(-\sigma\omega x_2, \sigma\omega x_1, 0, 0)$, where σ is of the form (11.3). In the next section we give a sketch of the main properties of such an electromagnetic field.

12. ELECTROMAGNETIC FIELD DUE TO 'CIRCULAR CURRENT'

In this section we shall not examine the internal and external fields in detail—in fact, we only discuss the periodic terms in the external field. If 'present' values are used, the magnetic and electric fields will be of the form

$$\begin{aligned} \mathbf{H} &= \mathbf{H}' \cos pt + \mathbf{H}'' \sin pt, \\ \mathbf{E} &= \mathbf{E}' \sin pt + \mathbf{E}'' \cos pt. \end{aligned}$$

To avoid unnecessary work we shall discuss fields of the simple form

$$\mathbf{H} = \mathbf{H}' \cos pt, \quad \mathbf{E} = \mathbf{E}' \sin pt.$$

For $\mathbf{E}$ to be of the order p^2 , we require the condition $\text{curl } \mathbf{H}'_0 = 0$, where $\mathbf{H}_0$ is the leading term in $\mathbf{H}$. This condition with $\nabla \mathbf{H}'_0 = 0$ severely limits the possible values of $\mathbf{H}'_0$. We start off our solution by taking

$$H'_{0n} = \left(\frac{x_n x_3}{r^5} - \frac{1}{3r^3} \delta_{n3} \right) p$$

and omit a constant factor throughout. The equations (10.1) then give for $\mathbf{E}'_0$, the leading term of $\mathbf{E}$,

$$E'_{0n} = \frac{1}{3r^3} (-x_2 \delta_{n1} + x_1 \delta_{n2}) p^2.$$

By successive approximation we obtain

$$H_n = \cos pt \left\{ \left(\frac{1}{r^5} p + \frac{1}{6r^3} p^3 + \frac{1}{24r} p^5 \right) x_n x_3 + \delta_{n3} \left(-\frac{1}{3r^3} p + \frac{1}{6r} p^3 - \frac{1}{8} r p^5 \right) \right\},$$

$$E_n = \sin pt (-x_2 \delta_{n1} + x_1 \delta_{n2}) \left(\frac{1}{3r^3} p^2 + \frac{1}{6r} p^4 \right).$$

By (10.3), the energy-tensor representing this field is

$$\left. \begin{aligned} 16\pi E_{mn} &= \lambda^2 x_m x_n \left(\frac{1}{r^4} - \frac{x_3^2}{r^6} \right) p^6, \\ 16\pi E_{44} &= \lambda^2 \left(\frac{1}{r^2} - \frac{x_3^2}{r^4} \right) p^6, \\ 16\pi E_{4n} &= 0, \end{aligned} \right\} \quad (12.1)$$

where λ^2 is a positive constant. Referring to (6.6) and (6.7), if we put

$$A_3 = -\frac{1}{18}, \quad B_2 = -\frac{2}{45}, \quad A'_2 = -\frac{43}{15}, \quad B'_1 = -\frac{14}{15}, \quad A_1 = \frac{29}{30},$$

the expressions for the energy-tensor take the forms

$$16\pi T_{mn} = -\frac{24}{5} I^2 p^6 x_m x_n \left(\frac{1}{r^4} - \frac{x_3^2}{r^6} \right),$$

$$16\pi T_{44} = -\frac{24}{5} I^2 p^6 \left(\frac{1}{r^2} - \frac{x_3^2}{r^4} \right),$$

which, except for the sign, are of the same form as the above values of E_{mn} , E_{44} .

If retarded and not present values are used, the fields are given by

$$H_n = \frac{x_n x_3}{r^5} \left\{ \left(p - \frac{1}{3} p^3 r^2 \right) \cos p(t-r) - p^2 r \sin p(t-r) \right\}$$

$$+ \frac{1}{r^3} \delta_{n3} \left\{ \left(-\frac{1}{3} p + \frac{1}{3} p^3 r^2 - \frac{1}{9} p^5 r^4 \right) \cos p(t-r) + \left(\frac{1}{3} p^2 r - \frac{1}{9} p^4 r^3 \right) \sin p(t-r) \right\}$$

$$+ \text{terms of order } p^6,$$

$$E_n = \frac{1}{3r^3} (-x_2 \delta_{n1} + x_1 \delta_{n2}) \left\{ p^2 \sin p(t-r) + \left(p^3 r - \frac{1}{3} p^5 r^3 \right) \sin p(t-r) \right\}$$

$$+ \text{terms of order } p^6.$$

The non-periodic parts of E_{44} , E_{mn} are the same as those given in (12.1). On the other hand, the partial contributions to E_{4n} are given by

$$E_1 H_3 = \left(-\frac{x_2}{18r^3} + \frac{x_2 x_3^2}{9r^5} \right) p^6, \quad E_2 H_3 = \left(\frac{x_1}{18r^3} - \frac{x_1 x_3^2}{9r^5} \right) p^6,$$

and

$$E_1 H_2 - E_2 H_1 = \left(\frac{x_3}{9r^3} - \frac{x_3^2}{9r^5} \right) p^6.$$

Thus we see that the momentum terms are not zero but, except for some positive factor, are of the form

$$4\pi E_{4n} = \left\{ \frac{1}{2} \left(-\frac{x_n}{r^3} + \frac{x_3}{r^3} \delta_{n3} \right) + \frac{x_n x_3^2}{r^5} \right\} p^6.$$

It is easily seen that it is not possible to find potentials γ_{4n} which satisfy the equation

$$\gamma_{4n,44} - \gamma_{44,4n} = E_{4n}.$$

In this way we find that, in general, the value of the non-periodic part of E_{4n} depends on whether 'present' or 'retarded' values are used, although the non-periodic parts of E_{mn} , E_{44} are the same.

In the next section we investigate the manner in which the theory develops in the case of a rotating sphere. Before doing this, it is helpful to summarize briefly the results which we have so far obtained.

In § 6 we found that it is impossible to obtain a solution of the equation $G'_{\mu\nu} = 0$ for a rotating nearly spherical body. In §§ 11 and 12 we have shown that the energy-tensor can be reduced to a form which represents an electromagnetic field. Furthermore, in general, gravitational potentials cannot be found so that the momentum components T_{4n} vanish; this is because we have used 'retarded' instead of 'present' values. The same phenomenon occurs in the corresponding electromagnetic field.

The reversal of the sign of the energy-tensor is rather curious but not new. A similar reversal occurs in the theory of gravitational mass (Clark 1949c, p. 422). Later researches on this aspect of the theory will no doubt shed light on Eddington's Dublin lectures (1943, §§ 11, 13) in which he discusses the reversal of sign by a partition of the energy-tensor into a 'particle' energy-tensor and a 'field' energy-tensor.*

13. RELATIVISTIC ELECTRODYNAMICS

When electrical phenomena are considered, an antisymmetrical covariant tensor, $F_{\mu\nu}$, is introduced in relativity theory. The components of $F_{\mu\nu}$ are given by the equations

$$E_n = F_{n4}, \quad H_n = \epsilon_{nik} F_{ik}, \quad (13.1)$$

where, as in (10.3), ϵ_{nik} is the Levi-Civita tensor.

In general theory Maxwell's equations (10.1) assume the forms

$$(F_{\mu\nu})_{;\sigma} = 0, \quad (13.2)$$

$$4\pi J^\mu = (F^{\mu\nu})_{;\nu}, \quad (13.3)$$

where () denotes covariant differentiation and J^μ the current-charge vector.

In the case of static fields (13.2) reduces to the ordinary classical equations

$$\text{div } \mathbf{H} = 0, \quad \text{curl } \mathbf{E} = 0. \quad (13.4)$$

We suppose that the leading (i.e. classical) terms in $F_{\mu\nu}$ are of the same order as the gravitational density; these terms will be denoted by $F_{\mu\nu}^{(1)}$. We also restrict the first-order terms of the gravitational potentials to be of the form

$$h_{mn} = \phi \delta_{mn}, \quad h_{44} = \phi, \quad h_{4n},$$

where ϕ and h_{4n} are given respectively by (9.2) and (8.1).

* The reversal of sign can be obtained by assuming that there is an electric polarization vector and a magnetization vector such that $\mathbf{D} = -\mathbf{E}$ and $\mathbf{B} = -\mathbf{H}$.

Under these conditions the expression (13.3) for the current-charge vector yields the results

$$\left. \begin{aligned} 4\pi J^m &= F_{ms|s} + \phi_{|s} F_{ms}^{(1)} - h_{4m|s} F_{4s}^{(1)} - h_{4s} F_{m4|s}^{(1)} \\ 4\pi J^4 &= -F_{4s|s} + \phi_{|s} F_{4s}^{(1)} - h_{4r|s} F_{rs}^{(1)} \end{aligned} \right\} \quad (13.5)$$

provided that terms of the order of the cube of the density are neglected and $F_{\mu s|s}^{(1)} = 0$.

We start by taking the leading terms

$$\left. \begin{aligned} H_n^{(1)} &= \frac{1}{5} m k_1 a^2 \omega \left(-\frac{1}{r^3} \delta_{n3} + \frac{3x_n x_3}{r^5} \right), \\ E_n^{(1)} &= \frac{m k_2 x_n}{r^3} + m k_3 a^4 \omega^2 \left(\frac{x_n}{r^5} + \frac{2x_3}{r^5} \delta_{n3} - \frac{5x_n x_3^2}{r^7} \right). \end{aligned} \right\} \quad (13.6)$$

Here m is simply the gravitational mass and the k_i pure numbers owing to the fact that we have adopted units such that the velocity of light and the gravitational constant are unity.

Next we add terms $H_n^{(2)}$, $E_n^{(2)}$ of order m^2 , where

$$\left. \begin{aligned} H_n^{(2)} &= m^2 \lambda_3 a^2 \omega \left(-\frac{1}{2r^4} \delta_{n3} + \frac{x_n x_3}{r^6} \right), \\ E_n^{(2)} &= -\frac{2m^2 k_2 \lambda_4 x_n}{r^4} + m^2 a^4 \omega^2 \left(\frac{\lambda_1 x_n}{r^6} + \frac{\lambda_2 x_3}{3r^6} \delta_{n3} - \frac{\lambda_2 x_n x_3^2}{r^8} \right). \end{aligned} \right\} \quad (13.7)$$

$$\text{The fields} \quad \mathbf{H} = \mathbf{H}^{(1)} + \mathbf{H}^{(2)}, \quad \mathbf{E} = \mathbf{E}^{(1)} + \mathbf{E}^{(2)} \quad (13.8)$$

satisfy the equations (13.4). Again, when ϕ has the value (9.2) and h_{4n} the value (8.1), the expressions (13.6) give a current-charge vector whose components are

$$\left. \begin{aligned} 4\pi J^m &= m^2 a^2 \omega (\lambda_3 - \frac{2}{5} k_1 + 3A k_2) (x_2 \delta_{n1} - x_1 \delta_{n2}) \frac{1}{r^6}, \\ 4\pi J^4 &= m^2 a^4 \omega^2 \left\{ \left(-3\lambda_1 + \frac{1}{3} \lambda_2 + \frac{1}{5} A k_1 + B_4 k_2 - 2k_3 \right) \frac{1}{r^6} + \left(\lambda_2 - \frac{3}{5} A k_1 - 3B_4 k_2 + 6k_3 \right) \frac{x_3^2}{r^8} \right\}. \end{aligned} \right\} \quad (13.9)$$

14. THE STRESS-ENERGY TENSOR

The classical expressions for the stress-energy tensor has already been given in § 10. We first substitute (13.8) in these expressions and denote the result by $E_{\mu\nu}^0$. It is convenient to denote the terms of order m^2 by $E_{\mu\nu}^{(1)}$.

When a gravitational field is also present, the electromagnetic stress-tensor is determined by the relation

$$4\pi E_{\mu}^{\sigma} = -F^{\sigma\alpha} F_{\mu\alpha} + \frac{1}{4} g_{\mu}^{\sigma} F^{\alpha\beta} F_{\alpha\beta},$$

which reduces to

$$\begin{aligned} 4\pi E_{mn} &= 4\pi E_{mn}^0 + \phi (2F_{ms}^{(1)} F_{ns}^{(1)} - \frac{1}{2} F_{rs}^{(1)} F_{rs}^{(1)} \delta_{mn} - 4\pi E_{mn}^{(1)}) \\ &\quad - h_{4r} F_{n4}^{(1)} F_{mr}^{(1)} - h_{4r} F_{m4}^{(1)} F_{nr}^{(1)} + \delta_{mn} 4\pi h_{4r} E_{r4}^{(1)}, \end{aligned} \quad (14.1)$$

when the potentials are of the form (9.2), (8.1) and terms of order m^4 are neglected.

Forming the divergence, we obtain after some calculation

$$\begin{aligned}
 16\pi E_{mn|n} = & -4m^2 k_2^2 \lambda_4 \frac{x_m}{r^7} + 16\pi (h_{4r} E_{r4}^{(1)})_{|m} + m^3 a^4 \omega^2 \\
 & \times \left\{ \frac{x_m}{r^9} \left(-\frac{3}{2} A k_1 k_2 + \frac{4}{5} k_1^2 + 8k_2 k_3 - 8k_2 k_3 \lambda_4 - 2B_4 k_2^2 \right. \right. \\
 & \quad \left. \left. + 12\lambda_1 k_2 - \frac{4}{3} \lambda_2 k_2 - \frac{4}{5} \lambda_3 k_1 \right) \right. \\
 & + \frac{x_3}{r^9} \delta_{m3} \left(-\frac{3}{5} A k_1 k_2 + \frac{1}{2} \frac{6}{5} k_1^2 + 16k_2 k_3 + 4B_4 k_2^2 - \frac{8}{5} k_1 \lambda_3 - 16k_2 k_3 \lambda_4 \right) \\
 & + \frac{x_3^2 x_m}{r^{11}} \left(\frac{4}{5} A k_1 k_2 - \frac{3}{2} \frac{6}{5} k_1^2 - 40k_2 k_3 + 2B_4 k_2^2 \right. \\
 & \quad \left. \left. - 4\lambda_2 k_2 + \frac{1}{5} k_1 \lambda_3 + 40k_2 k_3 \lambda_4 \right) \right\}. \quad (14.2)
 \end{aligned}$$

15. THE GRAVITATIONAL POTENTIALS

In § 9 we have found that we cannot find a solution of the equation $G_{\mu\nu} = 0$ for a rotating sphere. We were able to find, however, gravitational potentials such that

$$16\pi T_{mn|n} = m^3 a^4 \omega^2 \left(\frac{P' x_m}{r^9} + \frac{Q' x_3}{r^9} \delta_{m3} + \frac{R' x_3^2 x_m}{r^{11}} \right),$$

where P' , Q' and R' are non-zero numerical constants. We now suppose that the rotating sphere produces an electromagnetic field of the type discussed in §§ 13 and 14. In this case additional gravitational potentials $h_{\mu\nu}^{(e)}$ must be introduced to represent the energy-tensor $E_{\mu\nu}^{(1)}$. These are determined by the equation

$$M_{\mu\nu} = 16\pi E_{\mu\nu}^{(1)},$$

where the $M_{\mu\nu}$ are given by (2.2). In particular, we have

$$h_{44|ss}^{(e)} = 16\pi E_{44}^{(1)}, \quad h_{4n|ss}^{(e)} = 16\pi E_{4n}^{(1)}.$$

The extra terms $h_{\mu\nu}^{(e)}$ give rise to additional terms in $2\Lambda'_{mn}$ which we denote by $2\Lambda_{mn}^{(e)}$. The equations (3.2a) and (3.7) give

$$\begin{aligned}
 2\Lambda_{mn}^{(e)} = & -\phi h_{44|mss}^{(e)} - h_{4r}^{(1)} h_{4n|ssr}^{(e)} + h_{4r}^{(1)} h_{4r|mss}^{(e)} \\
 = & -\phi E_{44|m}^{(1)} - h_{4r}^{(1)} E_{4n|r}^{(1)} + h_{4r}^{(1)} E_{4r|m}^{(1)},
 \end{aligned}$$

since $h_{4r|r}^{(e)} = 0$. The equation (5.5) may then be written in the form

$$16\pi T_{mn|n} = m^3 a^4 \omega^2 \left(\frac{P' x_m}{r^9} + \frac{Q' x_3}{r^9} \delta_{m3} + \frac{R' x_3^2 x_m}{r^{11}} \right) - 2\Lambda_{mn}^{(e)} - 16\pi (\phi E_{44}^{(1)})_{|m} + 16\pi (h_{4r} E_{4r}^{(1)})_{|m}. \quad (15.1)$$

On working out the value of the right-hand member of (15.1) and comparing with (14.2), we find that (15.1) reduces to

$$16\pi T_{mn|n} = 16\pi E_{mn|n},$$

provided that the λ_i are given by

$$\lambda_1 = \frac{2}{3}B_4k_2 - \frac{4}{3}k_3 + \frac{k_1}{12k_2}Q + \frac{1}{36}R - \frac{1}{12}P,$$

$$\lambda_2 = -\frac{3}{5}Ak_1 + 3B_4k_2 - 6k_3 + \frac{3k_1}{8k_2}Q + \frac{R}{4},$$

$$\lambda_3 = \frac{2}{5}k_1 - 3Ak_2 + \frac{5}{8}Q,$$

$$\lambda_4 = 1,$$

and

$$P = -P', \quad Q = -Q', \quad R = -R'.$$

On the other hand, if the signs of the potentials $h_{\mu\nu}^{(e)}$ are reversed and $P = P'$, etc., the equation (15.1) reduces to

$$16\pi T_{mn|n} = -16\pi E_{mn|n}.$$

The investigation on periodic systems shows that the second alternative is the one which actually occurs in nature. Also since the analysis in this section is only valid if k_2 has a non-zero value, we have the further result that there must be an internal charge.

Inserting the above values of the λ_i in the expressions (13.9), we obtain for the current-charge vector at great distance from the sphere

$$4\pi J^m = \frac{5}{8}m^2a^2\omega Q(x_2\delta_{n1} - x_1\delta_{n2})\frac{1}{r^6},$$

$$4\pi J^4 = m^2a^4\omega^2\left\{-\left(\frac{k_1}{8k_2}Q - \frac{P}{4}\right)\frac{1}{r^6} + \left(\frac{3k_1}{8k_2}Q + \frac{R}{4}\right)\frac{x_3^2}{r^8}\right\}.$$

By (9.5), with $P = P'$, etc., these expressions may be written

$$4\pi J^m = \frac{125}{2}m^2a^2\omega B'_4(x_2\delta_{n1} - x_1\delta_{n2})\frac{1}{r^6},$$

$$4\pi J^4 = \frac{1}{2}m^2a^4\omega^2 B'_4\left\{\left(\frac{169}{3} - 25\frac{k_1}{k_2}\right)\frac{1}{r^6} + \left(75\frac{k_1}{k_2} - 219\right)\frac{x_3^2}{r^8}\right\}.$$

16. ALTERNATIVE FIELD

An alternative solution can be obtained in the case of a rotating sphere by assuming that the rotation produces either magnetic or electric charges of order $m\omega$. Mathematically, the analysis is essentially the same for both assumptions. The argument, however, does not seem to be capable of extension to cover the periodic fields discussed earlier in the investigation. Moreover, the additional terms introduced in the internal energy-tensor do not appear to be useful in developing a relativity theory of elasticity. Therefore, we shall not consider this solution. In a future investigation I hope to examine the possibility that the rotation produces charges proportional to the square root of the density.

17. CONCLUSION

In 1923 H. A. Wilson suggested that a mass in motion would give rise to a magnetic field. He later rejected the idea as a result of an experiment which showed that a body in translational motion did not give such a field. In recent years Blackett (1947) has reintroduced Wilson's hypothesis in the case of rotational motion, but the theory has not been generally accepted. The present investigation establishes Blackett's hypothesis that a rotating body produces an electromagnetic field. The analysis does not, however, *prove* that the actual electromagnetic field is necessarily of the form suggested by Blackett. The calculation to the order of approximation considered verifies that a possible field is one whose leading terms (in c.g.s. electromagnetic units) are due to an internal current-charge vector

$$(-k_1\omega x_2, k_1\omega x_1, 0, k_2)\lambda^{\frac{1}{2}}\rho/c,$$

where k_1 and k_2 are non-zero numerical constants, λ the constant of gravitation, c the velocity of light and ρ the gravitational density.

REFERENCES

- Blackett, P. M. S. 1947 *Nature*, **159**, 658.
 Clark, G. L. 1941 *Proc. Roy. Soc. A*, **177**, 227.
 Clark, G. L. 1946 *Mon. Not. R. Astr. Soc.* **106**, 457.
 Clark, G. L. 1947 *Proc. Camb. Phil. Soc.* **43**, 164.
 Clark, G. L. 1948 *Phil. Mag.* ser. 7, **39**, 747.
 Clark, G. L. 1949a *Proc. Roy. Soc. Edinb.* **62**, 434.
 Clark, G. L. 1949b *Proc. Camb. Phil. Soc.* **45**, 405.
 Clark, G. L. 1949c *Proc. Roy. Soc. Edinb.* **62**, 412.
 Eddington, A. S. 1924 *Mathematical theory of relativity*, 2nd ed. Cambridge University Press.
 Eddington, A. S. 1943 *The combination of relativity and quantum theory*. Dublin: Inst. for Advanced Studies.
 Einstein, A. & Infeld, L. 1940 *Ann. Math. Princeton, etc.*, **41**, 455.
 Einstein, A., Infeld, L. & Hoffmann, B. 1938 *Ann. Math. Princeton, etc.*, **39**, 65.

Note on the problem of a rotating mass of perfect fluid in relativity mechanics

BY G. L. CLARK, PH.D.

Department of Mathematics, Bedford College, University of London

(Communicated by Sir Edmund Whittaker, F.R.S.—Received 16 January 1950)

A mass of perfect fluid rotating about an axis is considered. The material stress-system is taken to be one of isotropic pressure. General expressions for the pressure and form of the free surface are assumed. It is shown that the equations of motion cannot be satisfied unless the rotation produces an electromagnetic field which gives rise to additional terms in the total energy-tensor.

1. INTRODUCTION

In the preceding paper (Clark 1950)*, by considering the external field, it has been shown that a rotating body produces an electromagnetic field. In § 1 it was mentioned that before the investigation was undertaken some preliminary researches in the theory of elasticity had convinced the author of the reality of this phenomenon. The question whether the same conclusion can be derived in a purely hydromechanical theory forms the subject-matter of this paper. The problem chosen for investigation is relatively simple. Is it possible to satisfy the equations of motion in the case of a mass of liquid which rotates under isotropic pressure, if terms of order of the cube of the density are retained?

For the sake of definiteness the density referred to local co-ordinates is taken to be $\alpha + p$, where α is a constant and p denotes pressure. This distribution of density is one in which the velocity of sound is equal to that of light. General expressions, each containing several disposable constants, are assumed for the pressure and form of the free surface. The components of the stress-energy tensor are then computed and the results substituted in the equations of motion. I have not been able to obtain a solution which satisfies the equations of motion. It is possible, however, to satisfy these equations if we assume that the rotation produces an electromagnetic field.

2. THE ENERGY-TENSOR

A mass of liquid in motion under isotropic pressure (p) is represented by an energy-tensor of the form

$$T^{\mu\nu} = M^{\mu\nu} - g^{\mu\nu}p, \quad (2.1)$$

where

$$\begin{aligned} M^{\mu\nu} &= (\rho_{00} + p) \frac{dx_\mu}{ds} \frac{dx_\nu}{ds} \\ &= (\rho_{00} + p) v_\mu v_\nu \left(\frac{ds}{dt} \right)^2, \end{aligned} \quad (2.2)$$

and $v_\mu = dx_\mu/dt$. The density ρ_{00} is the density referred to local co-ordinates. We consider the case in which $\rho_{00} = \alpha + p$, α being a constant.

* This will be referred to as paper I.

When the motion is one of rotation about the x_3 -axis,

$$v_1 = -\omega x_2, \quad v_2 = \omega x_1, \quad v_3 = 0, \quad v_4 = 1.$$

We treat the Newtonian potentials as of the first order, the angular velocity ω is then of order $\frac{1}{2}$, and accordingly the order of the components of $T^{\mu\nu}$ are:

$$T^{44}, \text{ order } 1; \quad T^{4n}, \text{ order } \frac{3}{2}; \quad T^{mn}, \text{ order } 2.$$

The order of the terms required to be retained in the $g_{\mu\nu}$ is found (§ 3) to be as follows:

$$g_{44}, \text{ order } 2; \quad g_{4n}, \text{ order } \frac{3}{2}; \quad g_{mn}, \text{ order } 1.$$

We write

$$g_{44} = 1 + V + U,$$

where V and U represent the terms of only 1 and 2 respectively.

We use the same notation as in I; in particular, the quantities $h_{\mu\nu}$, $\gamma_{\mu\nu}$ have the same meaning. As far as the potentials of order 1 are concerned, we may take

$$h_{mn} = V\delta_{mn}, \quad h_{44} = V, \quad h_{4n} = 0.$$

Under these conditions

$$\left(\frac{ds}{dt}\right)^2 = 1 + V - (r^2 - x_3^2)\omega^2. \quad (2.3)$$

3. THE EQUATIONS OF MOTION

The equation of motion $(T^\nu_m)_{,\nu} = 0$ may be written in the alternative forms

$$\left. \begin{aligned} \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x_\nu} (T^\nu_m \sqrt{-g}) - \{m\nu, \alpha\} T^\nu_\alpha &= 0, \\ \frac{\partial T^{m\nu}}{\partial x_\nu} + \{\alpha\nu, \nu\} T^{m\alpha} + \{\alpha\nu, m\} T^{\alpha\nu} &= 0. \end{aligned} \right\} \quad (3.1)$$

When the potentials are independent of the time, these equations reduce to

$$\frac{\partial T^{ms}}{\partial x_s} - 2V_{|s} T^{ms} + \frac{1}{2} V_{|m} T^{44} + (h_{4s|m} - h_{4m|s}) T^{4s} + \frac{1}{2} V_{|m} (1 + V) T^{44} + \frac{1}{2} U_{|m} T^{44} = 0, \quad (3.2)$$

provided terms of order 4 are neglected. Substituting (2.1) in (3.2) the equation takes the form

$$\begin{aligned} \frac{\partial M^{ms}}{\partial x_s} - 2V_{|s} M^{ms} + \frac{1}{2} V_{|m} M^{44} + (h_{4s|m} - h_{4m|s}) M^{4s} \\ + \frac{1}{2} V_{|m} M^{44} + \frac{1}{2} \alpha V V_{|m} + \frac{1}{2} \alpha U_{|m} + (1 + V) p_{|m} = 0. \end{aligned} \quad (3.3)$$

In deriving (3.3) we have assumed that p is of order 2 and consequently $M^{44} = \alpha$ to the first order of approximation.

4. NEWTONIAN APPROXIMATION

If terms of order 3 in (3.3) are neglected, the equation reduces to

$$\frac{\partial M^{ms}}{\partial x_s} + \frac{1}{2} \alpha V_{|m} + p_{|m} = 0, \quad (4.1)$$

where the non-vanishing components of M^{mn} are

$$M^{11} = \alpha\omega^2 x_2^2, \quad M^{12} = -\alpha\omega^2 x_1 x_2, \quad M^{22} = \alpha\omega^2 x_1^2.$$

The equation (4.1) may therefore be written in the form

$$\alpha\omega^2(-x_m + x_3\delta_{m3}) + \frac{1}{2}\alpha V_m + p_m = 0. \quad (4.2)$$

We now assume that the form of the free surface is given by

$$r^2 = a^2 - \epsilon x_3^2, \quad (4.3)$$

and terms of order ϵ^2 are neglected.

The potential V is -2 (Newtonian potential); that is, the internal and external values are

$$\left. \begin{aligned} V^i &= 4\pi\alpha\left\{-a^2 + \frac{1}{3}r^2 + \epsilon\left(\frac{1}{3}a^2 - \frac{1}{15}r^2 + \frac{1}{5}x_3^2\right)\right\}, \\ V^0 &= -\frac{8\pi\alpha}{3r}\left(1 - \frac{1}{2}\epsilon\right) + \frac{4}{5}\pi\alpha a^5\epsilon\left(-\frac{1}{3r^3} + \frac{x_3^2}{r^5}\right). \end{aligned} \right\} \quad (4.4)$$

Taking the internal potential we have

$$\frac{1}{2}\alpha V_m = 4\pi\alpha^2\left\{\frac{1}{3}x_m + \epsilon\left(-\frac{1}{15}x_m + \frac{1}{5}x_3\delta_{m3}\right)\right\}. \quad (4.5)$$

We assume that the pressure is of the form

$$p = \pi\alpha^2\{A_1(a^2 - r^2) + \epsilon[-A_1x_3^2 + A_2(a^2 - r^2)]\}. \quad (4.6)$$

Clearly this expression vanishes at the boundary (4.3). We have

$$p_m = \pi\alpha^2\{-2A_1x_m + \epsilon(-2A_1x_3\delta_{m3} - 2A_2x_m)\}. \quad (4.7)$$

Finally, we assume that $\omega^2 = \pi\alpha(D_1 + D_2\epsilon)$. (4.8)

We substitute (4.5), (4.7) and (4.8) in (4.2); we find that the left-hand side is of the form

$$\pi\alpha^2\{p_1x_m + D_1x_3\delta_{m3} + \epsilon(p_2x_m + p_3x_3\delta_{m3})\}, \quad (4.9)$$

where

$$\left. \begin{aligned} p_1 &= -D_1 + \frac{4}{3} - 2A_1, \\ p_2 &= -D_2 - \frac{4}{15} - 2A_2, \\ p_3 &= D_2 + \frac{4}{5} - 2A_1. \end{aligned} \right\} \quad (4.10)$$

For each coefficient to be zero, we require

$$D_1 = 0, \quad A_1 = \frac{2}{3}, \quad A_2 = -\frac{2}{5}, \quad D_2 = \frac{8}{15}.$$

Inserting these values in (4.6) and (4.8) we obtain the classical values

$$p_0 = \frac{2}{3}\pi\alpha^2\{a^2 - r^2 + \epsilon(-x_3^2 - \frac{3}{5}a^2 + \frac{3}{5}r^2)\}, \quad (4.11)$$

$$\omega^2 = \frac{8}{15}\pi\alpha\epsilon. \quad (4.12)$$

5. NEXT APPROXIMATION

When terms of the next order are considered, we assume that the free surface is of the form

$$r^2 = a^2 - \epsilon x_3^2 + \pi\alpha\epsilon(G_1a^4 + G_2a^2x_3^2 + G_3x_3^4), \quad (5.1)$$

where the G_i are constants which have to be determined. We further assume that the pressure is of the form

$$\begin{aligned} p &= p_0 + \pi^2\alpha^3\{-(Q_1 + Q_2)a^4 + Q_1a^2r^2 + Q_2r^4 \\ &\quad + \epsilon[(Q_3a^4 + Q_4a^2r^2 + Q_5a^2x_3^2 + (\frac{2}{3}G_1 - Q_3 - Q_4)r^4 \\ &\quad + (\frac{2}{3}G_2 + Q_1 + 2Q_2 - Q_5)r^2x_3^2 + \frac{2}{3}G_3x_3^4]\}, \end{aligned} \quad (5.2)$$

where the Q_i are also disposable constants. It will be noticed that the right-hand side of (5.2) vanishes at the boundary (5.1). On working out the value of the expression on the left-hand side of (3.3) we find that it is of the form

$$\pi^2 \alpha^3 \{ K_1 a^2 x_m + K_2 x_m r^2 + \epsilon [K_3 a^2 x_m + K_4 a^2 x_3 \delta_{m3} + K_5 x_m r^2 + K_6 x_3^3 \delta_{m3} + H_1 x_3 r^2 \delta_{m3} + H_2 x_m x_3^2] \}. \quad (5.3)$$

By a suitable choice of the disposable constants, we can make the K_i have any desired value. In particular, we can choose the constants so that the K_i vanish. On the other hand, we find that the difference $H_1 - H_2$ is independent of the values of the disposable constants and this difference is not zero. This means that our proposed solution cannot satisfy the equations of motion. I have not tried to 'touch up' the expressions (5.1) and (5.2), as such an attempt would, I believe, be futile. The analysis, however, is not presented as an alternative proof of the phenomenon established by the more rigorous treatment in paper I.

6. THE MOMENTUM TERMS

Since $h_{4n} T^{4m}$ is of the order $\alpha^2 \omega^2$, and ω^2 is of order $\alpha \epsilon$, we only require the value of h_{4n} for a sphere. This is

$$h_{4n} = 8\pi\alpha\omega(-\frac{1}{3}a^2 + \frac{1}{5}r^2)(x_2\delta_{n1} - x_1\delta_{n2}).$$

$$\text{Also, } T^{4n} = \alpha\omega(-x_2\delta_{n1} + x_1\delta_{n2}).$$

Using these values we find

$$\begin{aligned} (h_{4m|s} - h_{4s|m}) T^{4s} &= 16\pi^2 \alpha^2 \omega^2 (-\frac{1}{3}a^2 x_m + \frac{1}{3}a^2 x_3 \delta_{m3} + \frac{2}{5}x_m r^2 - \frac{1}{5}x_3 r^2 \delta_{m3} - \frac{1}{5}x_m x_3^2) \\ &= \frac{128}{15} \pi^2 \alpha^3 \epsilon (-\frac{1}{3}a^2 x_m + \frac{1}{3}a^2 x_3 \delta_{m3} + \frac{2}{5}x_m r^2 - \frac{1}{5}x_3 r^2 \delta_{m3} - \frac{1}{5}x_m x_3^2). \end{aligned}$$

It will be noticed that these terms do not give a contribution to the difference $H_1 - H_2$.

7. THE SECOND APPROXIMATION TO h_{44}

$$\text{We have } G_{44} = -8\pi(T_{44} - \frac{1}{2}g_{44}T).$$

If terms up to order 2 only are retained, this equation reduces to

$$G_{44} = -4\pi\{(1+2V)T^{44} + T^u\}, \quad (7.1)$$

where in the present case

$$T^{44} = \frac{(\alpha + 2p_0)}{1 - (r^2 - x_3^2)\omega^2 + V} - p_0,$$

$$T^u = \alpha\omega_0^2(r^2 - x_3^2) + 3p_0.$$

Inserting these values in (7.1), the equation takes the form

$$-2G_{44} = 8\pi\{\alpha + 2\alpha\omega_0^2(r^2 - x_3^2) + \alpha V + 4p_0\}. \quad (7.2)$$

Now, with the notation used in I,

$$-2G_{44} = h_{44|ss} - 2L_{44}, \quad (7.3)$$

where to the order of approximation required

$$2L_{44} = -VV_{|ss} + V_{|s}V_{|s}. \quad (7.4)$$

The equations (7.2), (7.3) and (7.4) lead to an equation of the form

$$h_{44|ss}^{(i)} = 8\pi\alpha + 16\pi\alpha^2\{E_1a^2 + E_2r^2 + \epsilon(E_3a^2 + E_4r^2 + E_5x_3^2)\}.$$

The most general form for h_{44}^i is $h_{44}^i = V^i + U^i$, where V^i is given by (4.4) and

$$\begin{aligned} U^i = 16\pi^2\alpha^2\{ & p_1a^4 + \frac{1}{6}E_1a^2r^2 + \frac{1}{20}E_2r^4 \\ & + \epsilon[p_2a^4 + p_3a^2r^2 + (-3p_3 + \frac{1}{2}E_3)a^2x_3^2 \\ & + (\frac{1}{20}E_4 - \frac{1}{140}E_5 + \frac{3}{35}p_4)r^4 + (\frac{1}{14}E_5 - \frac{6}{7}p_4)r^2x_3^2 + p_4x_3^4]\}, \end{aligned} \quad (7.5)$$

and the p_i are constants. The values of p_i are determined by considering the external potential h_{44}^0 as well as the internal potential h_{44}^i , but in the present paper we do not require their values. The contributions to H_1 and H_2 in (5.3) from the term $\frac{1}{2}\alpha U_{|m}$ in (3.3) are therefore equal, the actual value being

$$(\frac{1}{14}E_5 - \frac{6}{7}p_4).$$

Accordingly, there is no contribution to the difference $H_1 - H_2$ from this term.

Similarly, from the form of the expression (5.2) we see that there is no contribution to $H_1 - H_2$ from the term $p_{|m}$.

8. REMAINING TERMS

It is readily verified that the quantity $V_{|s}M^{ms}$ is identically zero and that the contributions to H_1 and H_2 from $\frac{1}{2}\alpha VV_{|m}$ are the same, their value each being $\frac{1}{15}\frac{6}{5}$. The contributions to $H_1 - H_2$ therefore arise from the terms

$$\frac{1}{2}V_{|m}M^u, \quad Vp_{|m} \quad \text{and} \quad \frac{1}{2}V_{|m}M^{44}.$$

We first show that the contributions from the first two of these terms cancel out.

We have

$$\begin{aligned} \frac{1}{2}V_{|m}M^u &= \frac{1}{2}V_{|m}\omega^2(r^2 - x_3^2) \\ &= \frac{3}{4}\frac{2}{5}\pi^2\alpha^3\epsilon(x_m r^2 - x_m x_3^2). \end{aligned}$$

That is, the contribution to $H_1 - H_2$ is $\frac{3}{4}\frac{2}{5}$. Again,

$$Vp_{|m} = 16\pi^2\alpha^3[\frac{1}{3}a^2x_m - \frac{1}{9}x_m r^2 + \epsilon\{-\frac{1}{15}a^2x_m + \frac{1}{3}a^2x_3\delta_{m3} + \frac{4}{45}x_m r^2 - \frac{1}{9}x_3 r^2\delta_{m3} - \frac{1}{15}x_m x_3^2\}].$$

The contributions to H_1 and H_2 from $Vp_{|m}$ are therefore $-\frac{1}{9}\frac{6}{5}$ and $-\frac{1}{15}\frac{6}{5}$ respectively. The contribution to $H_1 - H_2$ is therefore $-\frac{3}{4}\frac{2}{5}$ and cancels out with the contribution from $\frac{1}{2}V_{|m}M^u$.

We write

$$M^{44} = \frac{\alpha + (\mu + 1)p}{1 - (r^2 - x_3^2)\omega^2 + V},$$

where $\alpha + \mu p$ is the density referred to local co-ordinates and μ is assumed to be a numerical constant. It will be noted that for the velocity of sound in the material to be less than or equal to that of light we require the condition $|d\rho_{00}/dp| \leq 1$; that is, $|\mu| \leq 1$. To the order of approximation considered, we have

$$\frac{1}{2}V_{|m}M^{44} = \frac{1}{2}\alpha V_{|m} - \frac{1}{2}\alpha VV_{|m} + \frac{1}{2}(\mu + 1)pV_{|m} + \frac{1}{2}(r^2 - x_3^2)\omega^2V_{|m}.$$

The only contributions to $H_1 - H_2$ arise from the third and fourth terms on the right-hand side of the above equation. The contributions to H_1 and H_2 from the third term are respectively

$$H_1 = -\frac{8}{15}(\mu + 1), \quad H_2 = -\frac{8}{9}(\mu + 1).$$

That is, the contribution to $H_1 - H_2$ is

$$H_1 - H_2 = \frac{16}{45}(\mu + 1).$$

The contribution to $H_1 - H_2$ from the fourth term is the same as that from $\frac{1}{2}V_m M^n$; as we have seen it is $\frac{32}{45}$. Accordingly, the value of the difference $H_1 - H_2$ in (5.3) is $(\mu + 3)\frac{16}{45}$ and this cannot be zero unless $\mu = -3$. In particular, when $\mu = 1$, $H_1 - H_2 = \frac{64}{45}$.

9. ELECTROMAGNETIC FIELD

If we assume that the rotation produces an electromagnetic field which is the same as that due to a current-charge vector

$$(-k_1\alpha\omega x_2, k_1\alpha\omega x_1, 0, k_2\alpha),$$

and insert the additional stress-energy terms in (5.3), we find that we can obtain a solution so that all the coefficients in (5.3) are zero, provided $k_2 \neq 0$. The same conclusion is obtained if we assume that the electromagnetic field is due to a current-charge vector

$$(-k_2\alpha\omega x_2, k_2\alpha\omega x_1, 0, k_2\alpha),$$

together with the field due to a magnetized body whose intensity of magnetization is $I_0 + I_1(x_1^2 + x_2^2)$. The calculations are not particularly interesting, and are therefore omitted. The result that $k_2 \neq 0$ was also derived in I. In consequence of this fact, the discussion in § 4 for the terms of order 2 has to be modified.

For the internal electric field, we have

$$E_m = 4\pi k_2 \alpha \left\{ \frac{1}{3}x_m + \epsilon \left(-\frac{1}{15}x_m + \frac{1}{5}x_3 \delta_{m3} \right) \right\}.$$

Therefore $E_m k_2 \alpha = 4\pi k_2^2 \alpha^2 \left\{ \frac{1}{3}x_m + \epsilon \left(-\frac{1}{15}x_m + \frac{1}{5}x_3 \delta_{m3} \right) \right\}$.

If we assume a reversal of sign of the electromagnetic energy-tensor, the new values of the coefficients p_i in (4.8) are found to be

$$\begin{aligned} p_1 &= \frac{4}{3}(1 - k_2^2) - 2A_1, \\ p_2 &= -\frac{4}{15}(1 - k_2^2) - D_2 - 2A_2, \\ p_3 &= D_2 + \frac{4}{5}(1 - k_2^2) - 2A_1. \end{aligned}$$

The conditions $p_i = 0$ lead to the relations

$$A_1 = \frac{2}{3}(1 - k_2^2), \quad D_2 = \frac{8}{15}(1 - k_2^2), \quad A_2 = -\frac{2}{5}(1 - k_2^2).$$

That is, $p_0 = \frac{2}{3}\pi\alpha^2(1 - k_2^2)\{a^2 - r^2 + \epsilon(-x_3^2 - \frac{3}{5}a^2 + \frac{3}{5}r^2)\}$,
 $\omega_0^2 = \frac{8}{15}\pi\alpha\epsilon(1 - k_2^2).$

The calculation, at least to the order considered, does not determine the electromagnetic field precisely.

REFERENCE

Clark, G. L. 1950 *Proc. Roy. Soc. A*, **201**, 488.

Theoretical studies in nuclear structure

I. Enumeration and classification of the states arising from the filling of the nuclear d -shell

BY H. A. JAHN, *Department of Mathematics, University College, Southampton*

(Communicated by R. E. Peierls, F.R.S.—Received 12 October 1949)

Using a simple method due to Racah, the orbital states arising from the filling of the nuclear d -shell in Russell-Saunders coupling are enumerated and, again following Racah, a new classification of the states introduced according to their transformation properties under the group of rotations in the five-dimensional space of the orbital states of a single d -particle. This new classification acquires significance from the short-range nature of nuclear forces and permits of predictions relating to the order of the levels with attractive forces of short range.

INTRODUCTION

The levels arising from the filling of the nuclear p -shell, in the Hartree approximation, have been studied in detail by Hund (1937), Feenberg & Wigner (1937), Feenberg & Phillips (1937) and Racah (1942*a*) (see Rosenfeld (1948) for an account of this work). A start is made here for the corresponding problem of the nuclear d -shell.

We begin in § 1 with a determination of the charge-spin structure of all Wigner supermultiplets (Wigner 1937*a*) up to and including those for 10 particles, using a new simple method of calculation due to Racah. In § 2 a similar method is used in enumerating the orbital states arising from the d^n configuration. We start with the classification of the orbital states according to irreducible representations of the group U_5 of unitary transformations in the five-dimensional space of the orbital states of a single d -particle. As is well known this introduces simultaneously the classification of the states according to the partition numbers characterizing irreducible representations of the group of permutations. The classification of states of definite partition numbers into states of different total angular momentum L is then introduced by consideration of the group of three-dimensional rotations R_3 which is a subgroup of U_5 . This classification still leads to a large number of states with the same L and the same partition numbers, e.g. for d^{10} [442] (d -shell half-filled, most symmetrical partition), we have the states (table 11):

d^{10} [442]	S^6	P^5	D^{15}	F^{12}	G^{18}	H^{13}	I^{15}	K^9	L^9	M^4	N^4	O	Q
with	$L = 0$	1	2	3	4	5	6	7	8	9	10	11	12

These states when combined with charge-spin functions of adjoint symmetry ([3322] $\equiv$ [11]) give rise to ^{13}L and ^{31}L totally antisymmetric states (Wigner supermultiplets) with the above L values. Now by analogy with the p -shell one might think that these states would occur in the order of L with those of lowest L lowest (formula of Hund (1937) and Racah (1942*a*)). However, this is not the case, for detailed calculations made by the writer for the d^2 , d^3 , d^4 configurations (to be reported upon

later), using the Rosenfeld optimum central force mixture, show that the levels are by no means ordered according to L , and that states of high L may well be low lying. Thus without detailed calculation for given forces it would appear not to be possible to predict which of the above Wigner supermultiplets would be the lowest.

However, following a suggestion made to the writer by Professor Racah, we may make an additional classification of the orbital states of the d -shell by the device of considering a group G contained in U_5 and containing R_3 ($R_3 \subset G \subset U_5$) and such that the five states ($m_l = 2, 1, 0, -1, -2$) of a single d -particle spread out an irreducible representation of G . Such a group G is the group R_5 of rotations in five-dimensional space. If the reduction of the representation [442] of U_5 is taken in the stages $U_5 \rightarrow R_5 \rightarrow R_3$ we obtain the results shown in table 1 for d^{10} [442] (see § 3, tables 13 and 14).

TABLE 1. EXAMPLE OF THE NEW CLASSIFICATION OF THE ORBITAL STATES OF d^n

group		representation												total
		[442]												
U_5	$[f_1 f_2 f_3]$	(00)	(20)	(20)	(22)	(22)	(31)	(40)	(32)	(42)	(42)	(43)	(44)	
R_5	$(\lambda_1 \lambda_2)$													
R_3	$L=0(S)$	1	.	.	1	1	.	.	.	1	1	.	1	S^6
	$1(P)$	.	.	.	.	.	1	.	1	1	1	1	.	P^5
	$2(D)$	.	1	1	1	1	1	1	2	2	2	2	1	D^{15}
	$3(F)$	.	.	.	1	1	2	.	1	2	2	2	1	F^{12}
	$4(G)$	.	1	1	1	1	1	1	2	3	3	2	2	G^{18}
	$5(H)$	.	.	.	.	.	2	1	2	2	2	3	1	H^{13}
	$6(I)$	.	.	.	1	1	1	1	1	3	3	2	2	I^{15}
	$7(K)$	.	.	.	.	.	1	.	1	2	2	2	1	K^9
	$8(L)$	.	.	.	.	.	.	1	1	2	2	2	1	L^9
	$9(M)$	.	.	.	.	.	.	.	.	1	1	1	1	M^4
	$10(N)$	.	.	.	.	.	.	.	.	1	1	1	1	N^4
	$11(O)$	.	.	.	.	.	.	.	.	.	.	1	.	O
	$12(Q)$	.	.	.	.	.	.	.	.	.	.	.	1	Q

The irreducible representations of R_5 are characterized by two integral numbers $(\lambda_1 \lambda_2)$, and the representation [442] of U_5 reduces as follows into representations of R_5 :

[442] = (00) + 2 (20) + 2 (22) + (31) + (40) + (32) + 2 (42) + (43) + (44).

The columns of the table indicate the manner in which the representations $(\lambda_1 \lambda_2)$ of R_5 reduce into irreducible representations of R_3 (characterized by L). From the final column we see how the states of d^{10} [442] with the same L are distributed over the different representations of R_5 . Now it was shown in the case of the atomic d -shell by Laporte (1942) and Laporte & Platt (1942) that for special ratios of the Slater radial integrals certain states coincided, and it was shown further by Racah (1943) that this Laporte-Platt degeneracy was a coincidence of levels with the same seniority number. The same degeneracy occurs also in the nuclear case for the special Laporte-Platt ratios of the radial integrals and corresponds to coincidence of levels with the same partition numbers $[f_1 f_2 f_3]$ and the same $(\lambda_1 \lambda_2)$. Whereas, however, the Laporte-Platt ratios are far removed from the actual ratios in the atomic case owing to the long-range nature of the Coulomb forces, this is no longer the case in the nucleus

where the forces are of short range. It may be shown easily that the Laporte-Platt ratios hold exactly in the fictitious limit of δ -type interactions, in which two particles interact only when they coincide. The calculations mentioned above for the d^2, d^3, d^4 configurations using forces of reasonable range, show that in the nucleus one is not so far removed from the Laporte-Platt or δ -type interaction limit. These calculations showed a grouping together of levels with the same $[f_1 f_2 f_3]$ and $(\lambda_1 \lambda_2)$, the groups of levels being ordered by $(\lambda_1 \lambda_2)$ with the (00) or (10) state lowest. They showed further that the states of given $(\lambda_1 \lambda_2)$ were arranged in the order of decreasing L with the state of highest L lowest. Thus the states of $d^4[4]$ occur in the following order (using the standard notation $^{2T+1, 2S+1}L_J$):

$d^4[4]: (00) ^{11}S_0; (20) ^{11}G_4 ^{11}D_2; (40) ^{11}L_8 ^{11}I_6 ^{11}H_5 ^{11}G_4 ^{11}D_2,$

those of $d^3[3]$ in the order

$d^3[3]: (10) ^{22}D_{\frac{3}{2}, \frac{5}{2}}; (30) ^{22}I_{\frac{1}{2}, \frac{3}{2}} ^{22}G_{\frac{3}{2}, \frac{5}{2}} ^{22}F_{\frac{3}{2}, \frac{5}{2}} ^{22}S_{\frac{1}{2}},$

showing an interesting tendency to the condition for the occurrence of nuclear isomerism. Were it possible to generalize this result we would have a means of predicting, without detailed calculation, the lowest state arising from the various stages of the filling of the d -shell. The lowest supermultiplets according to this scheme are given in table 2 (from table 15). These results are summarized in terms of the number of protons and neutrons in the d -shell in table 3.

TABLE 2. PREDICTED GROUND STATES IN THE VARIOUS STAGES OF THE FILLING OF THE d -SHELL

	$T_\zeta=0$	$T_\zeta=1$	$T_\zeta=2$	$T_\zeta=3$	$T_\zeta=4$	$T_\zeta=5$
d^{20}	[0] (00) ^{11}S	.	.	.	.	.
$d^{18} d^2$	[2] (00) $^{13}_{31}S$	[2] (00) ^{31}S	.	.	.	.
$d^{16} d^4$	[4] (00) ^{11}S	[31] (11) $^{31}_{33}F$	[22] (00) ^{51}S	.	.	.
$d^{14} d^6$	[42] (00) $^{13}_{31}S$	[42] (00) ^{31}S	[321] (11) $^{51}_{53}F$	[222] (00) ^{71}S	.	.
$d^{12} d^8$	[44] (00) ^{11}S	[431] (11) $^{31}_{33}F$	[422] (00) ^{51}S	[3221] (11) $^{71}_{73}F$	[2222] (00) ^{91}S	.
d^{10}	[442] (00) $^{13}_{31}S$	[442] (00) ^{31}S	[4321] (11) $^{51}_{53}F$	[4222] (00) ^{71}S	[32221] (11) $^{91}_{93}F$	[22222] (00)
	$T_\zeta=\frac{1}{2}$	$T_\zeta=\frac{3}{2}$	$T_\zeta=\frac{5}{2}$	$T_\zeta=\frac{7}{2}$	$T_\zeta=\frac{9}{2}$	
$d^{19} d^1$	[1] (00) ^{22}D	.	.	.	.	
$d^{17} d^3$	[3] (10) ^{22}D	[21] (10) ^{42}D	.	.	.	
$d^{15} d^5$	[41] (10) ^{22}D	[32] (10) ^{42}D	[221] (10) ^{62}D	.	.	
$d^{13} d^7$	[43] (10) ^{22}D	[421] (10) ^{42}D	[322] (10) ^{62}D	[2221] (10) ^{82}D	.	
$d^{11} d^9$	[441] (10) ^{22}D	[432] (10) ^{42}D	[4221] (10) ^{62}D	[3222] (10) ^{82}D	[22221] (10) $^{10, 12}D$	

TABLE 3. GROUND STATES AS FUNCTION OF PARITY OF PROTON AND NEUTRON NUMBERS

no. of protons (P)	no. of neutrons (N)	lowest state
odd	even	} 2D_1 or 2D_1
even	odd	
even	even	1S_0
odd	odd	$\left\{ \begin{array}{l} ^1S_0 \text{ or } ^3S_1 \text{ if } N=P \\ ^1F_3 \text{ or } ^3F_{2,3,4} \text{ if } N \neq P \end{array} \right.$

It would appear therefore that the simple assumption of a short-range attraction leads automatically to the result that an even number of protons and neutrons in the d -shell have an $1S_0$ level as their lowest state, whilst an odd proton or neutron added to an even number of protons and neutrons leads to a $2D_{3/2}$ or $2D_{5/2}$ state as the lowest state. With an odd number of protons and an odd number of neutrons however, the lowest state may have $J = 0, 1, 2, 3$ or 4 . The connexion with the Schmidt (1937) model of the nucleus is apparent.

1. CHARGE-SPIN STRUCTURE OF ALL WIGNER SUPERMULTIPLETS UP TO AND INCLUDING THOSE FOR 10 PARTICLES

The four charge-spin states $\tau_z = \pm \frac{1}{2}$, $\sigma_z = \pm \frac{1}{2}$ for a single particle spread out an irreducible representation of the unitary group U_4 in four dimensions. The charge-spin functions for n particles spread out a representation which may be reduced into irreducible representations of U_4 . These representations are characterized by the partition of n into four integral parts

$$n = g_1 + g_2 + g_3 + g_4 \quad \text{with} \quad g_1 \geq g_2 \geq g_3 \geq g_4.$$

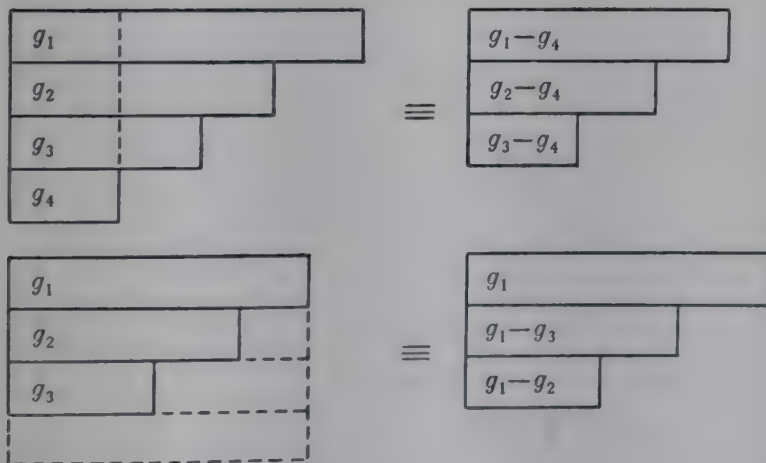
The representations

$$[g_1 g_2 g_3 g_4] \quad \text{and} \quad [g_1 - g_4, g_2 - g_4, g_3 - g_4, 0]$$

are equivalent, as are also the representations

$$[g_1 g_2 g_3 0] \quad \text{and} \quad [g_1, g_1 - g_3, g_1 - g_2, 0],$$

as is easily seen from the equivalence of the corresponding Young symmetry schemes,



The first equivalence was used by Wigner in his definition of the symbols (STY) or $(PP'P'')$ characterizing the representations

$$(S, T, Y) = \left(\frac{g'_1 + g'_2 - g'_3}{2}, \frac{g'_1 - g'_2 + g'_3}{2}, \frac{g'_1 - g'_2 - g'_3}{2} \right),$$

where $g'_1 = g_1 - g_4$, $g'_2 = g_2 - g_4$, $g'_3 = g_3 - g_4$ are the 'reduced' partition numbers. In terms of the four partition numbers $g_1 g_2 g_3 g_4$ the total dimension N of the corre-

sponding representation is given by the following formula (Weyl 1928, p. 248 or 1931, p. 336):

$$N[g_1g_2g_3g_4] = \frac{1}{12}(g_1 - g_2 + 1)(g_1 - g_3 + 2)(g_1 - g_4 + 3)(g_2 - g_3 + 1)(g_2 - g_4 + 2)(g_3 - g_4 + 1),$$

or by the equivalent formula

$$N[STY] = \frac{1}{12}(S + T + 3)(S - T + 1)\{(S + 2)^2 - Y^2\}\{(T + 1)^2 - Y^2\}.$$

The possible charge-spin supermultiplets up to and including those for $n = 10$ are listed in table 4 together with the dimension N .

The structure of a Wigner charge-spin supermultiplet, i.e. its constituent charge-spin TS multiplets, is given by the reduction of the corresponding representation of U_4 into irreducible representations of $U_2 \times U_2$. This reduction may be carried out

TABLE 4. DIMENSIONS OF IRREDUCIBLE REPRESENTATIONS OF U_4

n	partition	N	n	partition	N	n	partition	N
(0)	[0]	1	7	[7]	120		[54]	280
				[61]	216		[531]	360
1	[1]	4		[52]	224		[522]	160
				[511]	120		[5211] $\equiv$ [41]	
2	[2]	10		[43]	140		[441] $\equiv$ [43]	
	[11]	6		[421]	140		[432] $\equiv$ [421]	
				[331] $\equiv$ [32]			[4311] $\equiv$ [32]	
3	[3]	20		[4111] $\equiv$ [3]			[333] $\equiv$ [3]	
	[21]	20		[322] $\equiv$ [311]			[4221] $\equiv$ [311]	
	[111] $\equiv$ [1]			[3211] $\equiv$ [21]			[3321] $\equiv$ [21]	
				[2221] $\equiv$ [1]			[3222] $\equiv$ [1]	
4	[4]	35	8	[8]	165	10	[10]	280
	[31]	45		[71]	315		[91]	594
	[22]	20		[62]	360		[82]	770
	[211]	15		[611]	189		[811]	396
	[1111] $\equiv$ [0]			[53]	280		[73]	750
				[521]	256		[721]	640
				[5111] $\equiv$ [4]			[7111] $\equiv$ [6]	
5	[5]	56		[44]	105		[64]	540
	[41]	84		[431]	175		[631]	630
	[32]	60		[422]	84		[622]	270
	[311]	36		[4211] $\equiv$ [31]			[6211] $\equiv$ [51]	
	[221] $\equiv$ [21]			[332] $\equiv$ [31]			[55]	190
	[2111] $\equiv$ [1]			[3311] $\equiv$ [22]			[541]	384
				[3221] $\equiv$ [211]			[532]	300
6	[6]	84		[2222] $\equiv$ [0]			[5311] $\equiv$ [42]	
	[51]	140					[5221] $\equiv$ [411]	
	[42]	126	9	[9]	220		[442] $\equiv$ [42]	
	[411]	70		[81]	440		[4411] $\equiv$ [33]	
	[33]	50		[72]	540		[433] $\equiv$ [411]	
	[321]	64		[711]	280		[4321] $\equiv$ [321]	
	[222] $\equiv$ [2]			[63]	480		[3331] $\equiv$ [2]	
	[3111] $\equiv$ [2]			[621]	420		[4222] $\equiv$ [2]	
	[2211] $\equiv$ [11]			[6111] $\equiv$ [5]			[3322] $\equiv$ [11]	

(In the case of the d -shell we will be interested only in those partitions for which $g_1 \leq 5$.)

by a chain calculation of the following kind. We characterize the TS multiplets by their multiplicities $(2T+1, 2S+1)$ and start from the known structures given in table 5.

TABLE 5. STRUCTURE OF THE SIMPLEST CHARGE-SPIN SUPERMULTIPLETS

Young scheme	partition	$(2T+1, 2S+1)$	
$\begin{array}{ c } \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array}$	[0]	(11)	(singlet charge, singlet spin)
$\begin{array}{ c } \hline \square \\ \hline \end{array}$	[1]	(22)	(doublet charge, doublet spin)
$\begin{array}{ c c } \hline \square & \square \\ \hline \end{array}$	[2]	(11), (33)	(singlet charge, singlet spin) and (triplet charge, triplet spin)
$\begin{array}{ c c } \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}$	[11]	(13), (31)	(singlet charge, triplet spin) and (triplet charge, singlet spin)

To find the structure of the charge-spin supermultiplets for three particles we investigate the process of adding one further particle to two, i.e. the process of multiplying the charge-spin functions of two particles by those of a further particle. The permutation symmetry of the resulting states is obtained by adding a further square to the Young symmetry scheme in as many ways as possible as will give regular schemes (Littlewood 1940, theorem v, chapter 6, p. 94):

$$\begin{array}{c} \square \times \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array} + \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \\ \\ \square \times \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} = \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} + \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \equiv \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} + \begin{array}{|c|} \hline \square \\ \hline \end{array}, \end{array}$$

whilst the $(2T+1, 2S+1)$ structure of the resulting states is given by quantum vector addition. We thus obtain table 6. By subtracting the known structure of

TABLE 6. ADDITION OF ONE PARTICLE TO THE CHARGE-SPIN STATES OF TWO PARTICLES

$(2T+1, 2S+1)$	(22)	(24)	(42)	(44)	
[1](22) × [2](11)	1	.	.	.	} [3] + [21]
× [2](33)	1	1	1	1	
× [11](13)	1	1	.	.	} [21] + [111] ≡ [21] + [1]
× [11](31)	1	.	1	.	

[1] from that of [21] + [1] so obtained we derive immediately the structure of [21]; subtracting this in turn from that of [3] + [21] we obtain the structure of [3]:

$(2T+1, 2S+1)$:	(22)	(24)	(42)	(44)	
	2	1	1	0	[21] + [1]
	1	.	.	.	[1]
	1	1	1	0	[21]
	2	1	1	1	[3] + [21]
	1	0	0	1	[3]

Thus [21] has the $(2T+1, 2S+1)$ structure (22) + (24) + (42) with $N = 4 + 8 + 8 = 20$ and [3] has the structure (22) + (44) with $N = 4 + 16 = 20$.

Having found in this way the structure of all supermultiplets for $n = 3$ we find those for $n = 4$ and so on. As a further example we show in table 7 the calculation giving the structures for $n = 6$ knowing those for $n \leq 5$. In this table the abbreviation $\begin{pmatrix} 13 \\ 31 \end{pmatrix}$ for $(13) + (31)$ is used and at the end use is made of the fact that $[6]$ has the structure $(11) + (33) + (55) + (77)$, which is a special case of the general result that

$[2n + 1]$ has the structure $(22) + (44) + (66) + \dots + (2n + 2, 2n + 2)$

and $[2n]$ the structure $(11) + (33) + (55) + \dots + (2n + 1, 2n + 1)$.

The complete results of this calculation are contained in table 8.

TABLE 7. CHARGE-SPIN SUPERMULTIPLY STRUCTURES FOR
 $n = 6$ FROM THOSE FOR $n = 5$

$(2T + 1, 2S + 1):$	(11)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix}$	$\begin{pmatrix} 15 \\ 51 \end{pmatrix}$	$\begin{pmatrix} 17 \\ 71 \end{pmatrix}$	(33)	$\begin{pmatrix} 35 \\ 53 \end{pmatrix}$	$\begin{pmatrix} 37 \\ 73 \end{pmatrix}$	(55)	$\begin{pmatrix} 57 \\ 75 \end{pmatrix}$	(77)	
$[1](22) \times [5](22)$	1	1	.	.	1	.	.	.	.	.	} [6] + [51]
(44)	.	.	.	.	1	1	.	1	.	.	
(66)	.	.	.	.	.	.	.	1	1	1	
$\times [41](22)$	1	1	.	.	1	.	.	.	.	.	} [51] + [42] + [411]
$\begin{pmatrix} 24 \\ 42 \end{pmatrix}$	.	1	1	.	2	1	.	.	.	.	
(44)	.	.	.	.	1	1	.	1	.	.	
$\begin{pmatrix} 46 \\ 64 \end{pmatrix}$	.	.	.	.	.	1	1	2	1	.	
$\times [32](22)$	1	1	.	.	1	.	.	.	.	.	} [42] + [33] + [321]
$\begin{pmatrix} 24 \\ 42 \end{pmatrix}$	.	1	1	.	2	1	.	.	.	.	
$\begin{pmatrix} 26 \\ 62 \end{pmatrix}$	.	.	1	1	.	1	1	.	.	.	
(44)	.	.	.	.	1	1	.	1	.	.	
$\times [311](22)$	1	1	.	.	1	.	.	.	.	.	} [411] + [321] + [311] ≡ [411] + [321] + [111]
$\begin{pmatrix} 24 \\ 42 \end{pmatrix}$	.	1	1	.	2	1	.	.	.	.	
(44)	.	.	.	.	1	1	.	1	.	.	
$\times [221](22)$	1	1	.	.	1	.	.	.	.	.	} [321] + [222] + [221] ≡ [321] + [2] + [111]
$\begin{pmatrix} 24 \\ 42 \end{pmatrix}$	.	1	1	.	2	1	.	.	.	.	
$\times [2111](22)$	1	1	.	.	1	.	.	.	.	.	} [3111] + [2211] ≡ [2] + [111]
.	.	1	1	.	2	1	.	.	.	.	
1	1	1	.	.	3	1	.	.	.	.	[321]
.	1	.	.	.	1	1	.	1	.	.	[321] + [2]
1	1	1	1	1	2	2	1	1	.	.	[411]
1	1	1	.	.	3	2	1	2	1	.	[42] + [33]
1	1	.	.	.	2	1	.	2	1	1	[51] + [42]
1	.	.	.	.	1	.	.	1	.	1	[6] + [51]
.	1	.	.	.	1	1	.	1	1	.	[6]
1	.	1	.	.	2	1	1	1	.	.	[51]
.	1	.	1	.	.	1	.	.	.	.	[42]
.	1	.	1	1	.	1	.	.	.	.	[33]

TABLE 8. STRUCTURE OF ALL CHARGE-SPIN SUPERMULTIPLETS FOR $n \leq 10$

	partition	(STY)	(2T+1, 2S+1) structure	N
0)	[0]	(000)	(11)	1
1	[1]	($\frac{111}{2}$)	(22)	4
2	[2]	(111)	(11)(33)	10
	[11]	(100)	($\frac{13}{31}$)	6
3	[3]	($\frac{333}{2}$)	(22)(44)	20
	[21]	($\frac{311}{2}$)	(22)($\frac{24}{42}$)	20
4	[4]	(222)	(11)(33)(55)	35
	[31]	(211)	($\frac{13}{31}$)(33)($\frac{35}{53}$)	45
	[22]	(200)	(11)($\frac{15}{51}$)(33)	20
	[211]	(110)	($\frac{13}{31}$)(33)	15
5	[5]	($\frac{555}{2}$)	(22)(44)(66)	56
	[41]	($\frac{533}{2}$)	(22)($\frac{24}{42}$)(44)($\frac{46}{64}$)	84
	[32]	($\frac{511}{2}$)	(22)($\frac{24}{42}$)($\frac{26}{62}$)(44)	60
	[311]	($\frac{331}{2}$)	(22)($\frac{24}{42}$)(44)	36
6	[6]	(333)	(11)(33)(55)(77)	84
	[51]	(322)	($\frac{13}{31}$)(33)($\frac{35}{53}$)(55)($\frac{57}{75}$)	140
	[42]	(311)	(11)($\frac{15}{51}$)(33) ² ($\frac{35}{53}$)($\frac{37}{73}$)(55)	126
	[411]	(221)	($\frac{13}{31}$)(33)($\frac{35}{53}$)(55)	70
	[33]	(300)	($\frac{13}{31}$)($\frac{17}{71}$)($\frac{35}{53}$)	50
	[321]	(210)	($\frac{13}{31}$)($\frac{15}{51}$)(33) ² ($\frac{35}{53}$)	64
7	[7]	($\frac{777}{2}$)	(22)(44)(66)(88)	120
	[61]	($\frac{755}{2}$)	(22)($\frac{24}{42}$)(44)($\frac{46}{64}$)(66)($\frac{68}{86}$)	216
	[52]	($\frac{733}{2}$)	(22)($\frac{24}{42}$)($\frac{26}{62}$)(44) ² ($\frac{46}{64}$)($\frac{48}{84}$)(66)	224
	[511]	($\frac{553}{2}$)	(22)($\frac{24}{42}$)(44)($\frac{46}{64}$)(66)	120
	[43]	($\frac{711}{2}$)	(22)($\frac{24}{42}$)($\frac{26}{62}$)($\frac{28}{82}$)(44)($\frac{46}{64}$)	140
	[421]	($\frac{531}{2}$)	(22)($\frac{24}{42}$) ² ($\frac{26}{62}$)(44) ² ($\frac{46}{64}$)	140
8	[8]	(444)	(11)(33)(55)(77)(99)	165
	[71]	(433)	($\frac{13}{31}$)(33)($\frac{35}{53}$)(55)($\frac{57}{75}$)(77)($\frac{79}{97}$)	315
	[62]	(422)	(11)(33) ² (55) ² (77)($\frac{15}{51}$)($\frac{35}{53}$)($\frac{37}{73}$)($\frac{57}{75}$)($\frac{59}{95}$)	360
	[611]	(332)	(33)(55)(77)($\frac{13}{31}$)($\frac{35}{53}$)($\frac{57}{75}$)	189

TABLE 8 (cont.)

<i>n</i>	partition	(<i>STY</i>)	(2 <i>T</i> + 1, 2 <i>S</i> + 1) structure	<i>N</i>
8	[53]	(411)	(33)(55) $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 39 \\ 93 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$	280
	[521]	(321)	(33) ² (55) ² $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$	250
	[44]	(400)	(11)(33)(55) $\begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 19 \\ 91 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix}$	100
	[431]	(310)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} (33)^2 \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 37 \\ 73 \end{pmatrix} (55)$	170
	[422]	(220)	(11) $\begin{pmatrix} 15 \\ 51 \end{pmatrix} (33)^2 \begin{pmatrix} 35 \\ 53 \end{pmatrix} (55)$	80
9	[9]	$\begin{pmatrix} 9 & 9 & 9 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44)(66)(88)(10, 10)	220
	[81]	$\begin{pmatrix} 9 & 7 & 7 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44)(66)(88) $\begin{pmatrix} 24 \\ 42 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix} \begin{pmatrix} 68 \\ 86 \end{pmatrix} \begin{pmatrix} 8, 10 \\ 10, 8 \end{pmatrix}$	440
	[72]	$\begin{pmatrix} 9 & 5 & 5 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44) ² (66) ² (88) $\begin{pmatrix} 24 \\ 42 \end{pmatrix} \begin{pmatrix} 26 \\ 62 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix} \begin{pmatrix} 48 \\ 84 \end{pmatrix} \begin{pmatrix} 68 \\ 86 \end{pmatrix} \begin{pmatrix} 6, 10 \\ 10, 6 \end{pmatrix}$	540
	[711]	$\begin{pmatrix} 7 & 7 & 5 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44)(66)(88) $\begin{pmatrix} 24 \\ 42 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix} \begin{pmatrix} 68 \\ 86 \end{pmatrix}$	280
	[63]	$\begin{pmatrix} 9 & 3 & 3 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44) ² (66) $\begin{pmatrix} 24 \\ 42 \end{pmatrix} \begin{pmatrix} 26 \\ 62 \end{pmatrix} \begin{pmatrix} 28 \\ 82 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix}^2 \begin{pmatrix} 48 \\ 84 \end{pmatrix} \begin{pmatrix} 4, 10 \\ 10, 4 \end{pmatrix} \begin{pmatrix} 68 \\ 86 \end{pmatrix}$	480
	[621]	$\begin{pmatrix} 7 & 5 & 3 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44) ² (66) ² $\begin{pmatrix} 24 \\ 42 \end{pmatrix}^2 \begin{pmatrix} 26 \\ 62 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix}^2 \begin{pmatrix} 48 \\ 84 \end{pmatrix} \begin{pmatrix} 68 \\ 86 \end{pmatrix}$	420
	[54]	$\begin{pmatrix} 9 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44)(66) $\begin{pmatrix} 24 \\ 42 \end{pmatrix} \begin{pmatrix} 26 \\ 62 \end{pmatrix} \begin{pmatrix} 28 \\ 82 \end{pmatrix} \begin{pmatrix} 2, 10 \\ 10, 2 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix} \begin{pmatrix} 48 \\ 84 \end{pmatrix}$	280
	[531]	$\begin{pmatrix} 7 & 3 & 1 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44) ³ (66) $\begin{pmatrix} 24 \\ 42 \end{pmatrix}^2 \begin{pmatrix} 26 \\ 62 \end{pmatrix}^2 \begin{pmatrix} 28 \\ 82 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix}^2 \begin{pmatrix} 48 \\ 84 \end{pmatrix}$	360
	[522]	$\begin{pmatrix} 5 & 5 & 1 \\ 2 & 2 & 2 \end{pmatrix}$	(22)(44) ² (66) $\begin{pmatrix} 24 \\ 42 \end{pmatrix} \begin{pmatrix} 26 \\ 62 \end{pmatrix} \begin{pmatrix} 46 \\ 64 \end{pmatrix}$	160
10	[10]	(555)	(11)(33)(55)(77)(99)(11, 11)	280
	[91]	(544)	(33)(55)(77)(99) $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix} \begin{pmatrix} 79 \\ 97 \end{pmatrix} \begin{pmatrix} 9, 11 \\ 11, 9 \end{pmatrix}$	590
	[82]	(533)	(11)(33) ² (55) ² (77) ² (99) $\begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix} \begin{pmatrix} 59 \\ 95 \end{pmatrix} \begin{pmatrix} 79 \\ 97 \end{pmatrix} \begin{pmatrix} 7, 11 \\ 11, 7 \end{pmatrix}$	770
	[811]	(443)	(33)(55)(77)(99) $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix} \begin{pmatrix} 69 \\ 97 \end{pmatrix}$	390
	[73]	(522)	(33)(55) ² (77) $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 39 \\ 93 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}^2 \begin{pmatrix} 59 \\ 95 \end{pmatrix} \begin{pmatrix} 5, 11 \\ 11, 5 \end{pmatrix} \begin{pmatrix} 79 \\ 97 \end{pmatrix}$	750
	[721]	(432)	(33) ² (55) ² (77) ² $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}^2 \begin{pmatrix} 59 \\ 95 \end{pmatrix} \begin{pmatrix} 79 \\ 97 \end{pmatrix}$	640
	[64]	(511)	(11)(33) ² (55) ² (77) $\begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 19 \\ 91 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix}^2 \begin{pmatrix} 39 \\ 93 \end{pmatrix} \begin{pmatrix} 3, 11 \\ 11, 3 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix} \begin{pmatrix} 59 \\ 95 \end{pmatrix}$	540
	[631]	(421)	(33) ² (55) ³ (77) $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^3 \begin{pmatrix} 37 \\ 73 \end{pmatrix}^2 \begin{pmatrix} 39 \\ 93 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}^2 \begin{pmatrix} 59 \\ 95 \end{pmatrix}$	630
	[622]	(331)	(11)(33) ² (55) ² (77) $\begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$	270
	[55]	(600)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 1, 11 \\ 11, 1 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 39 \\ 93 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$	190
	[541]	(410)	(33) ² (55) ² $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 19 \\ 91 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 37 \\ 73 \end{pmatrix}^3 \begin{pmatrix} 39 \\ 93 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$	380
	[532]	(320)	(33) ² (55) ² $\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^3 \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$	300

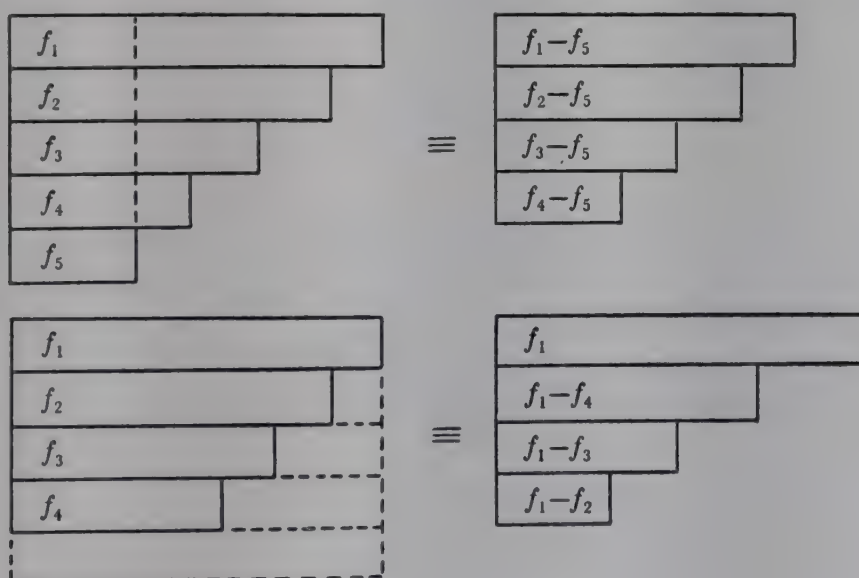
In table 8 the $(2T+1, 2S+1)$ structure is written as a product and multiples of the same TS multiplet indicated by indices. In each case it may be verified that the total dimension N agrees with the Weyl general formula. This extends calculations made by Wigner, and the method used is considerably simpler than that used by him. As mentioned in the Introduction the method was suggested to the writer by Professor Racah.

2. ENUMERATION OF THE ORBITAL STATES ARISING FROM d^n

The five states $m_l = 2, 1, 0, -1, -2$ of a single d -particle spread out an irreducible representation of the unitary group U_5 in five dimensions. The orbital functions for n d -particles spread out a representation which may be reduced into irreducible representations of U_5 . These representations are characterized by the partition of n into 5 integral parts:

$$n = f_1 + f_2 + f_3 + f_4 + f_5 \quad \text{with} \quad f_1 \geq f_2 \geq f_3 \geq f_4 \geq f_5.$$

From the Young symmetry schemes we see that we have the following equivalences:



or

$$[f_1 f_2 f_3 f_4 f_5] \equiv [f_1 - f_5, f_2 - f_5, f_3 - f_5, f_4 - f_5],$$

$$[f_1 f_2 f_3 f_4] \equiv [f_1, f_1 - f_4, f_1 - f_3, f_1 - f_2].$$

The dimension N of the representation $[f_1 f_2 f_3 f_4 f_5]$ is given by (Weyl 1928, p. 248, or 1931, p. 336)

$$N[f_1 f_2 f_3 f_4 f_5] = \frac{1}{2^8 8} (f_1 - f_2 + 1) (f_1 - f_3 + 2) (f_1 - f_4 + 3) (f_1 - f_5 + 4) \\ (f_2 - f_3 + 1) (f_2 - f_4 + 2) (f_2 - f_5 + 3) (f_3 - f_4 + 1) (f_3 - f_5 + 2) (f_4 - f_5 + 1).$$

Since the orbital states have to be combined with charge-spin states of adjoint symmetry ($[g_1 g_2 g_3 g_4] = [5f_5 4f_4 - f_5 3f_3 - f_4 2f_2 - f_3 1f_1 - f_2]$), it follows in the well-known way that only those representations with $f_1 \leq 4$ need be considered. The relevant representations and their dimensions are listed in table 9.

TABLE 9. DIMENSIONS OF IRREDUCIBLE REPRESENTATIONS OF U_5

n	partition	N	n	partition	N	n	partition	N
(0)	[0]	1		[3111]	70	9	[441]	980
1	[1]	5		[2211] $\equiv$ [211]			[432]	1120
2	[2]	15		[2111] $\equiv$ [1]			[4311]	720
	[11]	10	7	[43]	560		[333] $\equiv$ [33]	
3	[3]	35		[421]	700		[4221]	480
	[21]	40		[331]	315		[3321] $\equiv$ [321]	
	[111] $\equiv$ [11]			[4111]	160		[3222] $\equiv$ [3111]	
4	[4]	70		[322]	210		[42111] $\equiv$ [31]	
	[31]	105		[3211]	175		[33111] $\equiv$ [22]	
	[22]	50		[2221] $\equiv$ [21]			[32211] $\equiv$ [211]	
	[211]	45		[31111] $\equiv$ [2]		10	[22221] $\equiv$ [1]	
	[1111] $\equiv$ [1]			[22111] $\equiv$ [11]			[442]	1176
5	[41]	224	8	[44]	490		[4411]	700
	[32]	175		[431]	1050		[433] $\equiv$ [4411]	
	[311]	126		[422]	560		[4321]	1024
	[221]	75		[4211]	450		[3331] $\equiv$ [32]	
	[2111]	24		[332] $\equiv$ [331]			[4222]	200
	[11111] $\equiv$ [0]			[3311] $\equiv$ [322]			[3322] $\equiv$ [311]	
6	[42]	420		[3221] $\equiv$ [3211]			[43111] $\equiv$ [32]	
	[411]	280		[2222] $\equiv$ [2]			[42211] $\equiv$ [311]	
	[33]	175		[41111] $\equiv$ [3]			[33211] $\equiv$ [221]	
	[321]	280		[32111] $\equiv$ [21]			[32221] $\equiv$ [2111]	
	[222] $\equiv$ [22]			[22211] $\equiv$ [11]			[22222] $\equiv$ [0]	

The enumeration of the L -values occurring with given partition numbers is given by the reduction of the corresponding representation of U_5 into those of R_3 . This reduction may be carried out by a chain calculation of the same nature as that in § 1, making occasional use of the more general addition of two or more particles at a time. The mathematical theorem used is the following (Littlewood 1940, theorem v, chap. 6, p. 94):

To obtain the reduction of the product representation

$$[\mu_1\mu_2\ldots\mu_p][\lambda_1\lambda_2\ldots\lambda_q]$$

of the unitary group U_n we add to the $[\lambda]$ Young scheme

- μ_1 identical symbols α ,
- μ_2 identical symbols β ,
- μ_3 identical symbols γ ,
-

in this order, such that

- (1) a regular scheme is formed with no two identical symbols in the same column,
- (2) if all the added symbols are read from right to left in consecutive rows we obtain a lattice permutation of

$$\alpha^{\mu_1}\beta^{\mu_2}\gamma^{\mu_3}\ldots,$$

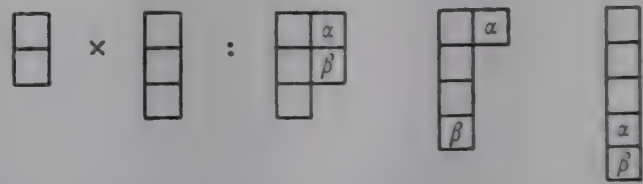
i.e. such that among the first r terms the number of times α occurs is *not less* than the number of times β occurs, similarly β does not occur fewer times than γ , etc., for all r .

TABLE 10. CHAIN CALCULATION OF THE L -STRUCTURE OF THE d^n CONFIGURATION

L	S	P	D	F	G	H	I	K	L	M	N	O	Q	n
	0	1	2	3	4	5	6	7	8	9	10	11	12	3
$[1] \times [2]$	1	1	3	2	2	1	1	.	.	.	.	.	$[3][21]$	3
$[1] \times [11]$	.	2	2	2	1	1	.	.	.	.	.	.	$[21][111] \equiv [21][11]$	3
$[1] \times [3]$	1	2	4	3	4	3	2	1	1	.	.	.	$[4][31]$	4
$[1] \times [21]$	2	4	5	6	5	3	2	1	.	.	.	.	$[31][22][211]$	4
$[1] \times [111]$	.	2	2	2	1	1	.	.	.	.	.	.	$[211][1111] \equiv [211][11]$	4
$[11] \times [2]$	.	4	3	5	3	3	1	1	.	.	.	.	$[31][211]$	4
$[1] \times [1111]$	1	1	1	1	1	.	.	.	.	.	.	.	$[2111][11111] \equiv [2111][0]$	5
$[11] \times [111]$	2	2	4	3	3	1	1	.	.	.	.	.	$[221][2111][0]$	5
$[1] \times [211]$	1	5	6	7	5	4	2	1	.	.	.	.	$[311][221][2111]$	5
$[1] \times [22]$	2	3	7	5	6	4	3	1	1	.	.	.	$[32][221]$	5
$[1] \times [31]$	2	7	9	11	10	9	6	4	2	1	.	.	$[41][32][311]$	5
$[1] \times [2111]$	1	3	4	4	3	2	1	.	.	.	.	.	$[3111][211][1]$	6
$[1] \times [221]$	3	6	9	9	9	6	4	2	1	.	.	.	$[321][22][211]$	6
$[1] \times [311]$	2	9	11	14	12	11	7	5	2	1	.	.	$[411][321][3111]$	6
$[11] \times [22]$	1	9	8	13	9	9	5	4	1	1	.	.	$[33][321][211]$	6
$[1] \times [32]$	4	9	14	16	16	13	11	7	4	2	1	.	$[42][33][321]$	6
$[11] \times [2111]$	2	5	7	7	6	4	2	1	.	.	.	.	$[3211][2][21][11]$	7
$[1] \times [3111]$	2	5	8	8	8	6	4	2	1	.	.	.	$[4111][3211][2]$	7
$[111] \times [22]$	1	9	8	13	9	9	5	4	1	1	.	.	$[331][3211][11]$	7
$[1] \times [33]$	1	9	11	14	13	14	9	8	5	3	1	1	$[43][331]$	7
$[111] \times [31]$	5	11	18	19	20	16	13	8	5	2	1	.	$[421][4111][3211][2]$	7
$[1] \times [321]$	6	16	23	27	26	22	17	11	6	3	1	.	$[421][331][322][3211]$	7
$[1] \times [3211]$	4	12	17	19	18	15	10	6	3	1	.	.	$[4211][322][3211][21]$	8
$[1] \times [322]$	5	11	18	19	20	16	13	8	5	2	1	.	$[422][331][3211]$	8
$[1] \times [331]$	4	16	21	27	25	24	18	14	8	5	2	1	$[431][331][322]$	8
$[11] \times [33]$	8	12	23	24	28	22	22	14	11	6	4	1	$[44][431][322]$	8
$[1] \times [332]$	4	16	21	27	25	24	18	14	8	5	2	1	$[432][33][321]$	9
$[1] \times [3311]$	5	11	18	19	20	16	13	8	5	2	1	.	$[4311][321][22]$	9
$[1] \times [3221]$	4	12	17	19	18	15	10	6	3	1	.	.	$[4221][321][3111][211]$	9
$[11] \times [331]$	12	25	43	46	51	43	38	26	19	10	6	2	$[441][432][4311][321][22]$	9
$[1] \times [333]$	1	9	11	14	13	14	9	8	5	3	1	1	$[433][32]$	10
$[1] \times [3321]$	6	16	23	27	26	22	17	11	6	3	1	.	$[4321][32][311][221]$	10
$[1] \times [3222]$	2	5	8	8	8	6	4	2	1	.	.	.	$[4222][311][2111]$	10
$[11] \times [332]$	12	25	43	46	51	43	38	26	19	10	6	2	$[442][433][4321][32][221]$	10

Then the product representation $[\mu][\lambda]$ contains representations corresponding to each such allowed scheme.

Examples for U_5



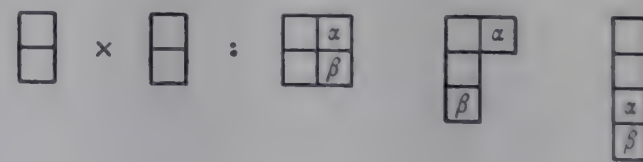
or $[11] \times [111] = [221] + [2111] + [11111] \equiv [221] + [2111] + [0].$

Dimension check: $10 \times 10 = 75 + 24 + 1.$

TABLE 10A. AUXILIARY VECTOR COUPLING TABLE

<i>L</i>	0	1	2	3	4	5	6	7	8	9	10	11	12	
$[1] \times (0)$	.	.	1	.	.	.	.	.	.	.	.	.	.	$(D) \times (L)$
(1)	.	1	1	1	.	.	.	.	.	.	.	.	.	
(2)	1	1	1	1	1	.	.	.	.	.	.	.	.	
(3)	.	1	1	1	1	1	.	.	.	.	.	.	.	
(4)	.	.	1	1	1	1	1	.	.	.	.	.	.	
(5)	.	.	.	1	1	1	1	1	.	.	.	.	.	
(6)	.	.	.	.	1	1	1	1	1	.	.	.	.	
(7)	.	.	.	.	.	1	1	1	1	1	.	.	.	
(8)	.	.	.	.	.	.	1	1	1	1	1	.	.	
(9)	.	.	.	.	.	.	.	1	1	1	1	1	.	
(10)	.	.	.	.	.	.	.	.	1	1	1	1	1	
(11)	.	.	.	.	.	.	.	.	.	1	1	1	1	
$[11] \text{ or } [111] \times (0)$	.	1	.	1	.	.	.	.	.	.	.	.	.	$(P + F) \times (L)$
(1)	1	1	2	1	1	.	.	.	.	.	.	.	.	
(2)	.	2	2	2	1	1	.	.	.	.	.	.	.	
(3)	1	1	2	2	2	1	1	.	.	.	.	.	.	
(4)	.	1	1	2	2	2	1	1	.	.	.	.	.	
(5)	.	.	1	1	2	2	2	1	1	.	.	.	.	
(6)	.	.	.	1	1	2	2	2	1	1	.	.	.	
(7)	.	.	.	.	1	1	2	2	2	1	1	.	.	
(8)	.	.	.	.	.	1	1	2	2	2	1	1	.	
(9)	.	.	.	.	.	.	1	1	2	2	2	1	1	
(10)	.	.	.	.	.	.	.	1	1	2	2	2	1	
(11)	.	.	.	.	.	.	.	.	1	1	2	2	2	

Note that although $[111] \equiv [11]$ in U_5 so far as dimension and structure is concerned we do not obtain the same result for $[11] \times [11]$:



or $[11] \times [11] = [22] + [211] + [1111] \equiv [22] + [211] + [1].$

Dimension check: $10 \times 10 = 50 + 45 + 5.$

The chain calculation is given in table 10, in deriving which use is made of the known L -structures of $[0]$, $[1]$, $[2]$ and $[11]$ and of the auxiliary table 10A which is just a vector-coupling table ($[1] = [D]$, $[11] \equiv [111] = (P) + (F)$). The resulting L -structure of the d^n configuration is given in table 11, the entries in this table, as they are made, being used in the next stage of the chain calculation. It may be verified easily in each case that the total dimension $N = \Sigma(2L + 1)$ given in the last column agrees with the general dimension formula. It is a rather remarkable result that d^{10} in the $[4411]$ and $[433]$ partitions (which are equivalent) has no S state.

TABLE 11. L -STRUCTURE OF THE d^n CONFIGURATION

n	L	S	P	D	F	G	H	I	K	L	M	N	O	Q	total dimension
0	$[f]$	0	1	2	3	4	5	6	7	8	9	10	11	12	1
0	$[0]$	1	.	.	.	.	.	.	.	.	.	.	.	.	1
1	$[1]$	.	.	1	.	.	.	.	.	.	.	.	.	.	5
2	$[2]$	1	.	1	.	1	.	.	.	.	.	.	.	.	15
2	$[11]$	.	1	.	1	.	.	.	.	.	.	.	.	.	10
3	$[3]$	1	.	1	1	1	.	1	.	.	.	.	.	.	35
3	$[21]$	.	1	2	1	1	1	.	.	.	.	.	.	.	40
4	$[4]$	1	.	2	.	2	1	1	.	1	.	.	.	.	70
4	$[31]$	.	2	2	3	2	2	1	1	.	.	.	.	.	105
4	$[22]$	2	.	2	1	2	.	1	.	.	.	.	.	.	50
4	$[211]$	.	2	1	2	1	1	.	.	.	.	.	.	.	45
5	$[41]$	1	2	3	4	4	3	3	2	1	1	.	.	.	224
5	$[32]$	1	2	4	3	4	3	2	1	1	.	.	.	.	175
5	$[311]$	.	3	2	4	2	3	1	1	.	.	.	.	.	126
5	$[221]$	1	1	3	2	2	1	1	.	.	.	.	.	.	75
5	$[2111]$	.	1	1	1	1	.	.	.	.	.	.	.	.	24
6	$[42]$	3	2	7	5	8	5	6	3	3	1	1	.	.	420
6	$[411]$	.	4	3	6	4	5	3	3	1	1	.	.	.	280
6	$[33]$	.	3	1	5	2	3	2	2	.	1	.	.	.	175
6	$[321]$	1	4	6	6	6	5	3	2	1	.	.	.	.	280
6	$[3111]$	1	1	2	2	2	1	1	.	.	.	.	.	.	70
7	$[43]$	1	4	7	7	8	8	6	5	4	2	1	1	.	560
7	$[421]$	3	6	10	11	12	10	9	6	4	2	1	.	.	700
7	$[331]$	.	5	4	7	5	6	3	3	1	1	.	.	.	315
7	$[4111]$	.	2	3	3	3	3	2	1	1	.	.	.	.	160
7	$[322]$	2	2	5	4	5	3	3	1	1	.	.	.	.	210
7	$[3211]$	1	3	4	5	4	3	2	1	.	.	.	.	.	175
8	$[44]$	4	1	6	4	8	4	7	3	4	2	2	.	1	490
8	$[431]$	2	9	12	16	15	15	12	10	6	4	2	1	.	1050
8	$[422]$	4	3	10	7	11	7	8	4	4	1	1	.	.	560
8	$[4211]$	1	6	6	9	8	8	5	4	2	1	.	.	.	450
9	$[441]$	4	5	11	11	14	11	12	8	7	4	3	1	1	980
9	$[432]$	3	9	14	16	17	16	13	10	7	4	2	1	.	1120
9	$[4311]$	2	7	10	12	12	11	9	6	4	2	1	.	.	720
9	$[4221]$	2	5	8	9	9	8	6	4	2	1	.	.	.	480
10	$[442]$	6	5	15	12	18	13	15	9	9	4	4	1	1	1176
10	$[4411]$	.	7	7	11	9	11	7	7	4	3	1	1	.	700
10	$[4321]$	4	10	14	18	18	15	13	9	5	3	1	.	.	1024
10	$[4222]$	2	1	5	3	5	3	3	1	1	.	.	.	.	200

It may be remarked in passing that the well-known structure of the nuclear p^n configuration can be derived very easily by this method (see table 12).

TABLE 12. STRUCTURE OF THE NUCLEAR p^n CONFIGURATION
BY THE NEW METHOD

		S	P	D	F	G	
L		0	1	2	3	4	
$[1] \times [11]$	$\equiv [1]$	1	1	1	.	.	$[21][0]$
$[1] \times [2]$		.	2	1	1	.	$[3][21]$
$[1] \times [21]$		1	2	2	1	.	$[31][2][1]$
$[1] \times [3]$		1	1	2	1	1	$[4][31]$
$[1] \times [31]$		1	2	3	2	1	$[41][31][2]$
$[1] \times [32]$	$\equiv [31]$	1	2	3	2	1	$[42][3][21]$
	$[0]$	1	.	.	.	.	
	$[1]$	.	1	.	.	.	
	$[2]$	1	.	1	.	.	
	$[3]$	.	1	.	1	.	
	$[21]$	.	1	1	.	.	
	$[4]$	1	.	1	.	1	
	$[31]$	.	1	1	1	.	
	$[41]$	.	1	1	1	1	
	$[42]$	1	.	2	1	1	

orbit		charge and spin			
p^n	$[f_1 f_2 f_3]$	L	$[g_1 g_2 g_3 g_4]$	(STY)	$(2T+1, 2S+1)$
p^0	$[0]$	S	$[0]$	(000)	(11)
p^1	$[1]$	P	$[1]$	$(\frac{1}{2}\frac{1}{2}\frac{1}{2})$	(22)
p^2	$[2]$	SD	$[11]$	(100)	$(13)(31)$
	$[11] \equiv [1]$	P	$[2]$	(111)	$(11)(33)$
p^3	$[3]$	PF	$[111] \equiv [1]$	$(\frac{1}{2}\frac{1}{2}-\frac{1}{2})$	(22)
	$[21]$	PD	$[21]$	$(\frac{3}{2}\frac{1}{2}\frac{1}{2})$	$(22)(24)(42)$
	$[111] \equiv [0]$	S	$[3]$	$(\frac{3}{2}\frac{3}{2}\frac{3}{2})$	$(22)(44)$
p^4	$[4]$	SDG	$[1111] \equiv [0]$	(000)	(11)
	$[31]$	PDF	$[211]$	(110)	$(13)(31)(33)$
	$[22] \equiv [2]$	SD	$[22]$	(200)	$(11)(15)(51)(33)$
	$[211] \equiv [1]$	P	$[31]$	(211)	$(13)(31)(33)(35)(53)$
p^5	$[41]$	$PDFG$	$[2111] \equiv [1]$	$(\frac{1}{2}\frac{1}{2}\frac{1}{2})$	(22)
	$[32] \equiv [31]$	PDF	$[221] \equiv [21]$	$(\frac{3}{2}\frac{1}{2}-\frac{1}{2})$	$(22)(24)(42)$
	$[311] \equiv [2]$	SD	$[311]$	$(\frac{3}{2}\frac{3}{2}\frac{1}{2})$	$(22)(24)(42)(44)$
	$[221] \equiv [1]$	P	$[32]$	$(\frac{5}{2}\frac{1}{2}\frac{1}{2})$	$(22)(24)(42)(26)(62)(44)$
p^6	$[42]$	SD^2FG	$[2211] \equiv [11]$	(100)	$(13)(31)$
	$[411] \equiv [3]$	PF	$[3111] \equiv [2]$	(111)	$(11)(33)$
	$[33] \equiv [3]$	PF	$[222] \equiv [2]$	$(11-1)$	$(11)(33)$
	$[321] \equiv [21]$	PD	$[321]$	(210)	$(13)(31)(33)^2(15)(51)(35)(53)$
	$[222] \equiv [0]$	S	$[33]$	(300)	$(13)(31)(35)(53)(17)(71)$

3. A NEW CLASSIFICATION OF THE ORBITAL STATES OF d^n

We derive now the new classification of the orbital states of d^n according to the representations of the orthogonal group R_5 in five dimensions. The irreducible representations of this group are characterized by two integers λ_1, λ_2 with $\lambda_1 \geq \lambda_2$. The dimension N of the representation $(\lambda_1 \lambda_2)$ is given by (Littlewood 1940, p. 236)

$$N(\lambda_1 \lambda_2) = \frac{1}{6}(\lambda_1 - \lambda_2 + 1)(\lambda_1 + \lambda_2 + 2)(2\lambda_1 + 3)(2\lambda_2 + 1).$$

We will be concerned only with those representations for which $\lambda_1 \leq 4$. The dimensions of these representations are given in the last column of table 14. The reduction of the representations $[\lambda]$ of U_5 into representations (η) of R_5 may be determined by a theorem of Littlewood's (Littlewood 1940, p. 240), viz.

$$[\lambda] = \sum_{(\eta)} g_{\delta\eta\lambda}(\eta),$$

where

$$[\delta][\eta] = \sum_{[\lambda]} g_{\delta\eta\lambda}[\lambda],$$

and

$$[\delta] = [0]; [2]; [4], [22]; [42], [222]; [44], [422], [2222]; [442], [4222], [22222], \dots,$$

i.e. partitions into even parts, where parts greater than 4 need not be considered in our case.

The product $[\delta][\eta]$ and hence the coefficients $g_{\delta\eta\lambda}$ are determined by the theorem used previously (Littlewood's theorem v). In this derivation use is made of the modification rules (Murnaghan 1938, p. 282) for representation symbols in non-standard form for R_5 , viz.

$$(\lambda_1 \lambda_2 \lambda_3) = 0 \quad \text{if} \quad \lambda_3 > 1,$$

$$(\lambda_1 \lambda_2 1) = \epsilon(\lambda_1 \lambda_2),$$

all $(\lambda_1 \lambda_2 \lambda_3 \lambda_4)$ zero except the following:

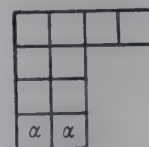
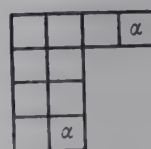
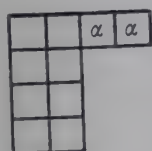
$$(\lambda_1 \lambda_2 22) = -(\lambda_1 \lambda_2),$$

$$(\lambda_1 \lambda_2 21) = -\epsilon(\lambda_1 \lambda_2),$$

$$(\lambda_1 111) = \epsilon(\lambda_1 0).$$

As an example of this method we give below the derivation of the reduction of the representation $[4222]$ of U_5 into irreducible representations of R_5 :

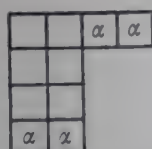
$$(a) [\delta] = [0]: (\eta) = (4222) \equiv -(42).$$



$$(b) [\delta] = [2]:$$

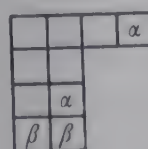
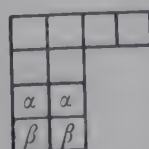
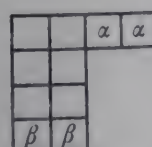
$$(\eta) = (2222) = -(22) \quad (\eta) = (3221) = -\epsilon(32) \quad \eta = (422) = 0$$

$$(c) [\delta] = [4]:$$



$$(\eta) = (222) = 0.$$

$$(d) [\delta] = [22]:$$



$$(\eta) = (222) = 0$$

$$(\eta) = (42) \quad \eta = (321) = \epsilon(32)$$

(e) $[\delta] = [42]$:

		α	α
α	α		
β	β		

$(\eta) = (22)$

(f) $[\delta] = [222]$:

		α	α
β	β		
γ	γ		

$(\eta) = (22)$

			α
	α		
β	β		
γ	γ		

$(\eta) = (31)$

α	α		
β	β		
γ	γ		

$(\eta) = (40)$

(g) $[\delta] = [422]$:

		α	α
α	α		
β	β		
γ	γ		

$(\eta) = (20)$

(h) $[\delta] = [2222]$:

		α	α
β	β		
γ	γ		
δ	δ		

$(\eta) = (20)$

(i) $[\delta] = [4222]$:

α	α	α	α
β	β		
γ	γ		
δ	δ		

$(\eta) = (00)$

Adding the results (a) to (i) we find, using the modification rules,

$$[4222] = (40) + (31) + (22) + 2(20) + (00).$$

Dimension check: $55 + 81 + 35 + 2(14) + 1 = 200 = N[4222].$

The reduction of the representations obtained in this manner is listed in table 13. In each case the sum of the dimensions checks with the total dimension.

Having obtained the reduction of the representations of U_5 both into those of R_3 (table 11) and into those of R_5 (table 13) it is a simple matter of subtraction to find the reduction of the representations of R_5 into those of R_3 . Thus for instance since we know that

$$[3] = (10) + (30) = SDFGI$$

and $[1] = (10) = D,$

we find $(30) = SFGI.$

The result for all the representations is given in table 14.

From the results so obtained we find the desired classification of the states arising from the filling of the nuclear d -shell given in table 15. This table contains the classification of both the orbital and the charge-spin parts of the totally antisymmetric wave functions.

TABLE 13. REDUCTION OF REPRESENTATIONS OF U_5 INTO REPRESENTATIONS OF R_5

n	U_5	R_5	N
0	[0]	(00)	1
1	[1]	(10)	5
2	[2]	(20) (00)	15
	[11]	(11)	10
3	[3]	(30) (10)	35
	[21]	(21) (10)	40
4	[4]	(40) (20) (00)	70
	[31]	(31) (20) (11)	105
	[22]	(22) (20) (00)	50
	[211]	(21) (11)	45
5	[41]	(41) (21) (30) (10)	224
	[32]	(32) (30) (21) (10)	175
	[311]	(31) (21) (11)	126
	[221]	(22) (21) (10)	75
	[2111]	(20) (11)	24
6	[42]	(42) (40) (31) (22) (20) ² (00)	420
	[411]	(41) (31) (21) (11)	280
	[33]	(33) (31) (11)	175
	[321]	(32) (31) (22) (21) (20) (11)	280
	[3111]	(30) (21) (10)	70
7	[43]	(43) (41) (32) (30) (21) (10)	560
	[421]	(42) (41) (32) (31) (22) (30) (21) ² (10)	700
	[331]	(33) (32) (31) (21) (11)	315
	[4111]	(40) (31) (20) (11)	160
	[322]	(32) (30) (22) (21) (10)	210
	[3211]	(31) (22) (21) (20) (11)	175
8	[44]	(44) (42) (40) (22) (20) (00)	490
	[431]	(43) (42) (33) (41) (32) (31) ² (22) (21) (20) (11)	1050
	[422]	(42) ((32) (40) (31) (22) ² (20) ² (00)	560
	[4211]	(41) (32) (31) (30) (21) ² (11)	450
9	[441]	(44) (43) (42) (41) (32) (22) (30) (21) (10)	980
	[432]	(43) (42) (41) (33) (32) ² (31) (30) (22) (21) ² (10)	1120
	[4311]	(42) (40) (33) (32) (31) ² (22) (21) (20) (11)	720
	[4221]	(41) (32) (31) (22) (30) (21) ² (10)	480
10	[442]	(44) (43) (42) ² (32) (40) (31) (22) ² (20) ² (00)	1176
	[4411]	(43) (33) (41) (32) (31) (21) (11)	700
	[4321]	(42) (41) (33) (32) ² (31) ² (30) (22) ² (21) ² (20) (11)	1024
	[4222]	(40) (31) (22) (20) ² (00)	200

TABLE 14. STRUCTURE AND DIMENSIONS OF REPRESENTATIONS OF R_5

	S	P	D	F	G	H	I	K	L	M	N	O	Q	dimension
$(\lambda_1 \lambda_2)$	0	1	2	3	4	5	6	7	8	9	10	11	12	N
(00)	1	.	.	.	.	.	.	.	.	.	.	.	.	1
(10)	.	.	1	.	.	.	.	.	.	.	.	.	.	5
(20)	.	.	1	.	1	.	.	.	.	.	.	.	.	14
(11)	.	1	.	1	.	.	.	.	.	.	.	.	.	10
(30)	1	.	.	1	1	.	1	.	.	.	.	.	.	30
(21)	.	1	1	1	1	1	.	.	1	.	.	.	.	35
(40)	.	.	1	.	1	1	1	1	.	.	.	.	.	55
(31)	.	1	1	2	1	2	1	1	.	.	.	.	.	81
(22)	1	.	1	1	1	.	1	.	.	.	.	.	.	35
(41)	.	1	1	2	2	2	2	2	1	1	.	.	.	154
(32)	.	1	2	1	2	2	1	1	1	.	.	.	.	105
(42)	1	1	2	2	3	2	3	2	2	1	1	.	.	220
(33)	.	1	.	2	1	1	1	1	.	1	.	.	.	84
(43)	.	1	2	2	2	3	2	2	2	1	1	1	.	231
(44)	1	.	1	1	2	1	2	1	1	1	1	.	1	165

TABLE 15. STATES ARISING FROM THE FILLING OF THE NUCLEAR *d*-SHELL

d^n	orbit				$[g_1 g_2 g_3 g_4]$	charge and spin			
	$[f_1 f_2 f_3 f_4]$	C	$(\lambda_1 \lambda_2)$	L		(STY)	$(2T+1, 2S+1)$		
d^0	[0]		(00)	S	[0]	(000)	(11)		
	[1]		(10)	D	[1]	$(\frac{1}{2}\frac{1}{2}\frac{1}{2})$	(22)		
d^2	[2]	7	(00)	S	[11]	(100)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix}$		
	[11]	2	(20)	GD	[2]	(111)	(11)(33)		
		0	(11)	FP					
d^3	[3]	13	(10)	D	[111] $\equiv$ [1]	$(\frac{1}{2}\frac{1}{2}-\frac{1}{2})$	(22)		
	[21]	6	(30)	$IGFS$					
		7	(10)	D				$(\frac{3}{2}\frac{1}{2}\frac{1}{2})$	$\begin{pmatrix} 24 \\ 42 \end{pmatrix}$
		3	(21)	$HGFDP$					
d^4	[111] $\equiv$ [11]	0	(11)	FP	[3]	$(\frac{3}{2}\frac{3}{2}\frac{3}{2})$	(22)(44)		
	[4]	26	(00)	S	[1111] $\equiv$ [0]	(000)	(11)		
	[31]	21	(20)	GD					
		12	(40)	$LIHGD$					
	[22]	15	(11)	FP	[211]	(110)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix}$ (33)		
		13	(20)	GD					
		8	(31)	KIH^2GF^2DP					
14		(00)	S						
[211]	9	(20)	GD	[22]	(200)	(11)(33) $\begin{pmatrix} 15 \\ 51 \end{pmatrix}$			
	6	(22)	$IGFDS$						
	7	(11)	FP				(211)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix}$ (33) $\begin{pmatrix} 35 \\ 53 \end{pmatrix}$	
4	(21)	$HGFDP$							
[1111] $\equiv$ [1]	0	(10)	D	[4]	(222)	(11)(33)(55)			

d^5	[41]	28	(10)	D	[2111] $\equiv$ [1]	(111)	(22)
		24	(21)	HGFDP			
		21	(30)	IGFS			
		15	(41)	MLK ³ I ² H ² G ² F ² DP			
	[32]	22	(10)	D	[221] $\equiv$ [21]	(21 - 1)	(22) $\begin{pmatrix} 24 \\ 42 \end{pmatrix}$
		18	(21)	HGFDP			
		15	(30)	IGFS			
		12	(32)	LKIH ² G ² FD ² P			
	[311]	17	(11)	FP	[311]	(331)	(22) $\begin{pmatrix} 24 \\ 42 \end{pmatrix}$ (44)
		14	(21)	HGFDP			
		10	(31)	KIH ² GF ² DP			
	[221]	14	(10)	D	[32]	(511)	(22) $\begin{pmatrix} 24 \\ 42 \end{pmatrix}$ (44) $\begin{pmatrix} 26 \\ 62 \end{pmatrix}$
d^6		10	(21)	HGFDP			
		8	(22)	IGFDS			
	[2111]	7	(11)	FP	[41]	(533)	(22) $\begin{pmatrix} 24 \\ 42 \end{pmatrix}$ (44) $\begin{pmatrix} 46 \\ 64 \end{pmatrix}$
		5	(20)	GD			
	[11111] $\equiv$ [0]	0	(00)	S	[5]	(555)	(22) (44) (66)
	[42]	37	(00)	S	[2211] $\equiv$ [11]	(100)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix}$
		32	(20) ²	(GD) ²			
		29	(22)	IGFDS			
		27	(31)	KIH ² GF ² DP			
		23	(40)	LIHGD			
		20	(42)	NML ² K ² I ³ H ² G ³ F ² D ² PS			
	[411]	30	(11)	FP	[3111] $\equiv$ [2]	(111)	(11) (33)
		27	(21)	HGFDP			
		23	(31)	KIH ² GF ² DP			
		18	(41)	MLK ² I ² H ² G ² F ² DP			
	[33]	30	(11)	FP	[222] $\equiv$ [2]	(11 - 1)	(11) (33)
		23	(31)	KIH ² GF ² DP			
		18	(33)	MKIHGF ² P			

TABLE 15 (cont.)

d^n	orbit			charge and spin		
	$[f_1 f_2 f_3 f_4]$	C	$(\lambda_1 \lambda_2)$	L	(STY)	$(2T+1, 2S+1)$
d^6 (cont.)	[321]	24	(11)	FP	(210)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 51 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}$
		22	(20)	GD		
		21	(21)	HGFDP		
		19	(22)	IGFDS		
		17	(31)	KIH ² GF ² DP		
	[222] $\equiv$ [22]	15	(32)	LKIH ² G ² FD ² P		
		21	(00)	S	(300)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix}$
		19	(20)	GD		
		13	(22)	IGFDS		
		19	(10)	D	(221)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 53 \end{pmatrix} \begin{pmatrix} 35 \\ 55 \end{pmatrix}$
d^7	[2211] $\equiv$ [211]	15	(21)	HGFDP		
		12	(30)	IGFS		
		14	(11)	FP	(311)	$(11) \begin{pmatrix} 33 \\ 51 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 55 \\ 53 \end{pmatrix} \begin{pmatrix} 37 \\ 3 \end{pmatrix}$
		11	(21)	HGFDP		
		7	(10)	D	(322)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 53 \end{pmatrix} \begin{pmatrix} 35 \\ 55 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$
	[21111] $\equiv$ [1]	45	(10)	D	$(\frac{1}{2}\frac{1}{2}-\frac{1}{2})$	(22)
		41	(21)	HGFDP		
		38	(30)	IGFS		
		35	(32)	LKIH ² G ² FD ² P		
		32	(41)	MLK ³ I ³ H ³ G ² F ³ DP		
	[43]	27	(43)	ONML ² K ² I ³ H ³ G ² F ³ D ² P		
		39	(10)	D		
		35	(21) ²	(HGFDP) ³		
		33	(22)	IGFDS		
		32	(30)	IGFS		
	[421]	31	(31)	KIH ² GF ² DP		
		29	(32)	LKIH ² G ² FD ² P		
		26	(41)	MLK ³ I ² H ² G ² F ² DP		
		24	(42)	NML ² K ² I ³ H ² G ³ F ² D ² PS		
					[3211] $\equiv$ [21]	$(\frac{3}{2}\frac{1}{2}\frac{1}{2}) \begin{pmatrix} 24 \\ 42 \end{pmatrix}$

d^7 (<i>cont.</i>)	[331]	34 31 27 25 22	(11) (21) (31) (32) (33)	FP HGFDP KIH ² GF ² DP LKI ² H ² G ³ FD ³ P MKIHGF ² P	[322] ≡ [311]	($\frac{3}{2}\frac{3}{2} - \frac{1}{2}$)	(22)($\frac{24}{42}$)(44)
	[4111]	32 30 25 21	(11) (20) (31) (40)	FP GD KIH ² GF ² DP LIHGD	[4111] ≡ [3]	($\frac{3}{2}\frac{3}{2}\frac{3}{2}$)	(22)(44)
	[322]	31 27 25 24 21	(10) (21) (22) (30) (32)	D HGFDP IGFDS IGFS LKI ² H ² G ² FD ² P	[331] ≡ [32]	($\frac{5}{2}\frac{1}{2} - \frac{1}{2}$)	(22)($\frac{24}{42}$)(44)($\frac{26}{62}$)
	[3211]	26 24 23 21 19	(11) (20) (21) (22) (31)	FP GD HGFDP IGFDS KIH ² GF ² DP	[421]	($\frac{5}{2}\frac{3}{2}\frac{1}{2}$)	(22)($\frac{24}{42}$) ² (44) ² ($\frac{26}{62}$)($\frac{46}{64}$)
	[2221] ≡ [21]	21 17	(10) (21)	D HGFDP	[43]	($\frac{7}{2}\frac{1}{2}\frac{1}{2}$)	(22)($\frac{24}{42}$)(44)($\frac{26}{62}$)($\frac{46}{64}$)($\frac{28}{82}$)
	[31111] ≡ [2]	21 16	(00) (20)	S GD	[511]	($\frac{5}{2}\frac{5}{2}\frac{3}{2}$)	(22)($\frac{24}{42}$)(44)($\frac{46}{64}$)(66)
	[22111] ≡ [11]	14	(11)	FP	[52]	($\frac{7}{2}\frac{3}{2}\frac{3}{2}$)	(22)($\frac{24}{42}$)(44) ² ($\frac{26}{62}$)($\frac{46}{64}$)(66)($\frac{48}{84}$)

TABLE 15 (cont.)

d^n	orbit	charge and spin						
		$[f_1 f_2 f_3 f_4]$	C	$(\lambda_1 \lambda_2)$	L	$[g_1 g_2 g_3 g_4]$	(STY)	$(2T+1, 2S+1)$
d^8	[44]		60	(00)	S	[2222] $\equiv$ [0]	(000)	(11)
			55	(20)	GD			
			52	(22)	$IGFDS$			
			46	(40)	$LIHGD$			
			43	(42)	$NML^2 K^2 I^3 H^2 G^3 F^2 D^2 PS$			
			36	(44)	$QNMLKI^2 HG^2 FDS$			
	[431]		49	(11)	FP	[3221] $\equiv$ [211]	(110)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} (33)$
			47	(20)	GD			
			46	(21)	$HGFDP$			
			44	(22)	$IGFDS$			
			42	(31) ²	$(KIH^2 GF^2 DP)^2$			
			40	(32)	$LKIH^2 G^2 FD^2 P$			
			37	(41)	$MLK^2 I^2 H^2 G^3 F^2 DP$			
			37	(33)	$MKIHGF^2 P$			
			35	(42)	$NML^2 K^2 I^3 H^2 G^3 F^2 D^2 PS$			
			32	(43)	$ONML^2 K^2 I^2 H^3 G^2 F^2 D^2 P$			
	[422]		48	(00)	S	[3311] $\equiv$ [22]	(200)	$(11) (33) \begin{pmatrix} 15 \\ 51 \end{pmatrix}$
			43	(20) ²	$(GD)^2$			
			40	(22) ²	$(IGFDS)^2$			
			38	(31)	$KIH^2 GF^2 DP$			
			34	(40)	$LIHGD$			
			36	(32)	$LKIH^2 G^2 FD^2 P$			
			31	(42)	$NML^2 K^2 I^3 H^2 G^3 F^2 D^2 PS$			
	[4211]		41	(11)	FP	[4211] $\equiv$ [31]	(211)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} (33) \begin{pmatrix} 35 \\ 53 \end{pmatrix}$
			38	(21) ²	$(HGFDP)^2$			
			35	(30)	$IGFS$			
			34	(31)	$KIH^2 GF^2 DP$			
			32	(32)	$LKIH^2 G^2 FD^2 P$			
			29	(41)	$MLK^2 I^2 H^2 G^2 F^2 DP$			

d^4 (cont.)	[332] $\equiv$ [331]	41 38 34 32 29	(11) (1) (31) (32) (33)	FP HGFDP KIH ² GF ² DP LKIH ² G ³ FD ³ P MKIHGF ² P	[332] $\equiv$ [31]	(21-1)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 31 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}$
	[3311] $\equiv$ [322]	38 34 32 31 28	(10) (21) (22) (30) (32)	D HGFDP IGFDS IGFS LKIH ² G ³ FD ³ P	[422]	(220)	$(11) \begin{pmatrix} 33 \\ 31 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 55 \\ 53 \end{pmatrix}$
	[3221] $\equiv$ [3211]	33 31 30 28 26	(11) (20) (21) (22) (31)	FP GD HGFDP IGFDS KIH ² GF ² DP	[431]	(310)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 31 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 37 \\ 71 \end{pmatrix} \begin{pmatrix} 37 \\ 71 \end{pmatrix}$
	[2222] $\equiv$ [2]	28 23	(00) (20)	S GD	[44]	(400)	$(11) \begin{pmatrix} 33 \\ 31 \end{pmatrix} \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 55 \\ 53 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 19 \\ 91 \end{pmatrix}$
	[41111] $\equiv$ [3]	34 27	(10) (30)	D IGFS	[5111] $\equiv$ [4]	(222)	$(11) \begin{pmatrix} 33 \\ 31 \end{pmatrix} \begin{pmatrix} 55 \\ 53 \end{pmatrix}$
	[32111] $\equiv$ [21]	28 24	(10) (21)	D HGFDP	[521]	(321)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 31 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$
	[22211] $\equiv$ [11]	21	(11)	FP	[53]	(411)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 31 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix} \begin{pmatrix} 39 \\ 93 \end{pmatrix}$
d^5	[441]	64 60 58 57 54 51 49 46 42	(10) (21) (22) (30) (32) (41) (42) (43) (44)	D HGFDP IGFDS IGFS LKIH ² G ² FD ² P MLK ² I ² H ² G ² F ² DP NML ² K ² I ³ H ² G ³ F ² D ² PS ONML ² K ² I ² H ³ G ³ F ² D ² P QNMLKI ² HG ² FDS	[3222] $\equiv$ [1]	$\begin{pmatrix} 11 \\ 11 \end{pmatrix}$	(22)

TABLE 15 (*cont.*)

d^n	orbit			charge and spin		
	$[f_1 f_2 f_3 f_4]$	C	$(\lambda_1 \lambda_2)$	L	$[g_1 g_2 g_3 g_4]$	(STY)
d^9 (<i>cont.</i>)	[432]	58	(10)	D	[3321] $\equiv$ [21]	$(2T+1, 2S+1)$
		54	(21) ²	$(HGFD P)^2$		$(22) \begin{pmatrix} 24 \\ 42 \end{pmatrix}$
		52	(22)	$IGFDS$		$(\frac{3}{2} \frac{1}{2} - \frac{1}{2})$
		51	(30)	$IGFS$		
		50	(31)	KIH^2GF^2DP		
		48	(32) ²	$(LKI H^2G^2FD^2P)^2$		
		45	(33)	$MKI HGF^2P$		
		45	(41)	$MLK^2I^2H^2G^2F^2DP$		
		43	(42)	$NML^2K^2I^3H^2G^3F^2D^2PS$		
		40	(43)	$ONML^2K^2I^2H^3G^2F^2D^2P$		
	[4311]	53	(11)	FP	[4221] $\equiv$ [311]	$(\frac{3}{2} \frac{3}{2} \frac{1}{2})$
		51	(20)	GD		$(22) \begin{pmatrix} 24 \\ 42 \end{pmatrix} (44)$
		50	(21)	$HGFD P$		
		48	(22)	$IGFDS$		
		46	(31) ²	$(KIH^2GF^2DP)^2$		
		44	(32)	$LKI H^2G^2FD^2P$		
		42	(40)	$LIHGD$		
		41	(33)	$MKI HGF^2P$		
		39	(42)	$NML^2K^2I^3H^2G^3F^2D^2PS$		
		51	(11)	FP	[333] $\equiv$ [3]	$(22) (44)$
	[333] $\equiv$ [33]	44	(31)	KIH^2GF^2DP		
		39	(33)	$MKI HGF^2P$		
		50	(10)	D	[4311] $\equiv$ [32]	$(\frac{5}{2} \frac{1}{2} \frac{1}{2})$
		46	(21) ²	$(HGFD P)^2$		$(22) \begin{pmatrix} 24 \\ 42 \end{pmatrix} (44) \begin{pmatrix} 26 \\ 62 \end{pmatrix}$
		44	(22)	$IGFDS$		
		43	(30)	$IGFS$		
		42	(31)	KIH^2GF^2DP		
		40	(32)	$LKI H^2G^2FD^2P$		
		37	(41)	$MLK^2I^2H^2G^2F^2DP$		

d^9 (<i>cont.</i>)	$[3321] \equiv [321]$	45 43 42 40 38 36	(11) (20) (21) (22) (31) (32)	FP GD HGFD IGFDS KIH ² GF ² DP LKI ² H ² G ² FD ² P	$[432] \equiv [421]$	$(\frac{53}{2}, \frac{1}{2})$	$(22) \binom{24}{42}^2 (44)^2 \binom{26}{62} \binom{46}{64}$
	$[3222] \equiv [3111]$	40 36 33	(10) (21) (30)	D HGFD IGFS	$[441] \equiv [43]$	$(\frac{71}{2}, -\frac{1}{2})$	$(22) \binom{24}{42} \binom{24}{42} \binom{26}{62} \binom{46}{64} \binom{28}{82}$
	$[42111] \equiv [31]$	43 41 36	(11) (20) (31)	FP GD KIH ² GF ² DP	$[5211] \equiv [41]$	$(\frac{533}{2}, \frac{3}{2})$	$(22) \binom{24}{42} \binom{24}{42} \binom{44}{64} \binom{46}{64}$
	$[33111] \equiv [22]$	42 37 34	(00) (20) (22)	S GD IGFDS	$[522]$	$(\frac{551}{2}, \frac{1}{2})$	$(22) \binom{24}{42} \binom{24}{42} \binom{44}{62}^2 \binom{46}{64} \binom{66}{64}$
	$[32211] \equiv [211]$	35 32	(11) (21)	FP HGFD	$[531]$	$(\frac{731}{2}, \frac{3}{2})$	$(22) \binom{24}{42}^2 \binom{44}{62}^3 \binom{26}{62}^2 \binom{46}{64}^2 \binom{28}{82} \binom{48}{84}$
	$[22221] \equiv [1]$	28	(10)	D	$[54]$	$(\frac{911}{2}, \frac{1}{2})$	$(22) \binom{24}{42} \binom{24}{42} \binom{44}{62} \binom{26}{62} \binom{46}{64} \binom{28}{82} \binom{48}{84} \binom{2,10}{10,2}$
d^{10}	$[442]$	75 70 67 65 63 61 58 55 51	(00) (20) ² (22) ² (31) (32) (40) (42) ² (43) (44)	S (GD) ² (IGFDS) ² KIH ² GF ² DP LKI ² H ² G ² FD ² P LIHGD (NML ² K ² I ³ H ² G ³ F ² D ² PS) ² ONML ² K ² I ² H ³ G ² F ² D ² P QNMLKI ² HG ² FDS	$[3322] \equiv [11]$	(100)	$\binom{13}{31}$

d^{10} (cont.)	[4222]	59	(00)	S	[4411] $\equiv$ [33]	(300)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix}$
		54	(20) ²	$(GD)^2$			
		51	(22)	$IGFDS$			
		49	(31)	KIH^2GF^2DP			
		45	(40)	$LIHGD$			
	[3322] $\equiv$ [311]	52	(11)	FP	[442] $\equiv$ [42]	(31-1)	$\begin{pmatrix} 11 \end{pmatrix} \begin{pmatrix} 33 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 55 \\ 73 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix}$
		49	(21)	$HGFDP$			
		45	(31)	KIH^2GF^2DP			
	[43111] $\equiv$ [32]	57	(10)	D	[5221] $\equiv$ [411]	(221)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \\ 53 \end{pmatrix} \begin{pmatrix} 35 \\ 55 \end{pmatrix}$
		53	(21)	$HGFDP$			
		50	(30)	$IGFS$			
		47	(32)	$LKIH^2G^3FD^2P$			
	[42211] $\equiv$ [311]	52	(11)	FP	[5311] $\equiv$ [42]	(311)	$\begin{pmatrix} 11 \end{pmatrix} \begin{pmatrix} 33 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 55 \\ 73 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix}$
		49	(21)	$HGFDP$			
		45	(31)	KIH^2GF^2DP			
	[33211] $\equiv$ [221]	49	(10)	D	[532]	(320)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^3 \begin{pmatrix} 55 \\ 71 \end{pmatrix}^2 \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix}$
		45	(21)	$HGFDP$			
		43	(22)	$IGFDS$			
	[32221] $\equiv$ [2111]	42	(11)	FP	[541]	(410)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 33 \end{pmatrix}^2 \begin{pmatrix} 15 \\ 51 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix}^2 \begin{pmatrix} 55 \\ 71 \end{pmatrix}^2 \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 37 \\ 73 \end{pmatrix}^2 \begin{pmatrix} 57 \\ 75 \end{pmatrix} \begin{pmatrix} 19 \\ 91 \end{pmatrix} \begin{pmatrix} 39 \\ 93 \end{pmatrix}$
		40	(20)	GD			
	[22222] $\equiv$ [0]	35	(00)	S	[55]	(500)	$\begin{pmatrix} 13 \\ 31 \end{pmatrix} \begin{pmatrix} 35 \\ 53 \end{pmatrix} \begin{pmatrix} 17 \\ 71 \end{pmatrix} \begin{pmatrix} 57 \\ 75 \end{pmatrix} \begin{pmatrix} 39 \\ 93 \end{pmatrix} \begin{pmatrix} 1, 11 \\ 11, 1 \end{pmatrix}$

Note added in proof. In table 15 has been added under the heading *C* the value of an expression which determines the order of the supermultiplets when only Wigner and Majorana forces of short range act: viz.

$$\frac{n(n+3)}{2} + 2(\alpha - \beta) - \frac{\lambda_1(\lambda_1+3) + \lambda_2(\lambda_2+1)}{2},$$

where α and β are the symmetry symbols of Hund (1937). A derivation of this expression, which involves the eigenvalues of Casimir's operator (cf. Racah 1949) for the group R_5 , will be given in a subsequent paper.

CONCLUSION

This paper brings the preliminary group-theoretical classification of the energy levels arising from the filling of the nuclear *d*-shell. Further papers will bring detailed calculations of the energy levels for definite assumed forces between the particles. As stated in the Introduction, however, it would appear that one can already make predictions about the order of levels without detailed calculation. Before detailed comparison with experiment can be made, detailed investigation will have to be made both of interconfigurational mixing and of the effect of non-central forces.

This work was undertaken in the main when the writer held a Research Fellowship in the Department of Mathematical Physics, Birmingham University. The writer is indebted to the University of Birmingham and to Professor R. E. Peierls for facilities afforded to carry out the research and to Professor G. Racah for informing him of unpublished work.

REFERENCES

- Feenberg, E. & Phillips, M. 1937 *Phys. Rev.* **51**, 597.
 Feenberg, E. & Wigner, E. 1937 *Phys. Rev.* **51**, 95.
 Hund, F. 1937 *Z. Phys.* **105**, 202.
 Laporte, O. 1942 *Phys. Rev.* **61**, 302.
 Laporte, O. & Platt, J. R. 1942 *Phys. Rev.* **61**, 305.
 Littlewood, D. 1940 *The theory of group characters*. Oxford: Clarendon Press.
 Murnaghan, F. D. 1938 *The theory of group representations*. Baltimore: The Johns Hopkins Press.
 Racah, G. 1942a *Phys. Rev.* **61**, 186.
 Racah, G. 1942b *Phys. Rev.* **62**, 438.
 Racah, G. 1943 *Phys. Rev.* **63**, 367.
 Racah, G. 1949 *Phys. Rev.* **76**, 1352.
 Rosenfeld, L. 1948 *Nuclear forces*. Amsterdam: North Holland Publishing Co.
 Schmidt, T. 1937 *Z. Phys.* **106**, 358.
 Weyl, H. 1928, 1931 *Gruppentheorie und Quantenmechanik*. Leipzig: Hirzel Verlag.
 Wigner, E. 1937a *Phys. Rev.* **51**, 106.
 Wigner, E. 1937b *Phys. Rev.* **51**, 947.
 Wigner, E. & Feenberg, E. 1941 *Rep. Progr. Phys.* **8**, 274.

The electrical conductivity of thin wires

By R. B. DINGLE, *The Royal Society Mond Laboratory, University of Cambridge*

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 12 October 1949—
Revised 26 November 1949)

In the first part of this paper, simple approximate methods have been developed for evaluating the electrical conductivity of films and wires of a size comparable with the mean free path of the conduction electrons.

In the second part, a rigorous theory has been given of the electrical conductivity of a thin wire, on the assumptions that the Fermi velocity surface is spherical and that the collisions of the electrons at the surface of the wire are inelastic. In the third part of the paper, this theory has been generalized to cover the case where the scattering is no longer inelastic.

In the final part, Andrew's recent experimental results for a thin mercury wire have been fitted to the theoretical curves obtained, and the mean free path evaluated.

1. INTRODUCTION

In metals at normal temperatures the mean free path of the conduction electrons is very much smaller than the least dimension of the specimen, and hence the electrical conductivity is independent of its size and shape. At low temperatures, however, owing to the decreased thermal vibration of the lattice, the mean free path is much increased, and the condition given above is no longer satisfied for thin films and wires.

If the electrons are specularly reflected at the internal surface of the metal, the conductivity remains the same as that of large specimens, since the momentum gained by an electron in the direction of the field is not lost on collision with the surface. If, however, the collisions at the surface are inelastic, the conductivity will fall owing to the increased transfer of momentum (directed along the field) from the electrons to the lattice. It is therefore necessary to introduce into the theory a parameter ϵ giving the probability that an electron is elastically scattered at the surface.

Previous theoretical work in this field may be divided into two categories, approximate and exact. Of the first, the most important is the method introduced by Lovell (1936), based on the simplifying assumption that all free paths start at the surface. This is a reasonable supposition if the specimen is thin compared with the mean free path, as then the number of collisions in the interior is much smaller than the number of collisions with the surface. This method has been applied by Lovell (1936) to the case of thin films, and by Andrew (1949) to that of thin wires.

The only exact work so far published is that of Fuchs (1938), who applied the methods of the electron theory of metals to the problem of the electrical conductivity of a thin film, thus obtaining an expression valid over the entire range of thicknesses.

The present paper is divided into four parts. In the first a simple argument is developed giving the conductivity of a specimen of any shape when the dimensions are considerably larger than the mean free path of the conduction electrons. In addition, the approximate theories for thin films and wires based on Lovell's assumptions are improved.

In the next two parts a rigorous solution will be found of the problem of the electrical conductivity of a wire, the considerations being based on the electron theory of metals. The essential assumption involved is that the Fermi velocity surface is spherical. The second part of the paper is concerned with the case of inelastic scattering, and the third part to the generalization to cover the case of partially elastic scattering. In the final part the experimental results recently obtained by Andrew (1949) for a thin mercury wire are compared with the theory.

It may be shown that the Wiedemann-Franz law relating the thermal and electrical conductivities holds to the first order in the temperature whatever the form of the energy surfaces or the degree of anisotropy of the relaxation time, provided only that such a time can be defined. Subject to this condition, the electronic contribution to the thermal conductivity will be reduced in small specimens by just the same ratio as the electrical conductivity.

PART I. APPROXIMATE METHODS

As the rigorous theory developed later in this paper is rather complicated, some simple approximate methods of obtaining the required results will first be discussed.

2. APPROXIMATE METHOD FOR THICK WIRES WITH CROSS-SECTIONS OF ARBITRARY SHAPE

When the wire is thick as compared to the mean free path, the deviation in the current density from that which would exist in the bulk metal for the same applied field will be appreciable only in a region near the surface of the specimen. The deviation in the current will be proportional to the surface area, whilst the mean current will be proportional to the volume. Let σ be the mean conductivity of the specimen, and σ_0 that of the bulk metal; we take the cross-section to be of perimeter P and area S , and λ to be the mean free path of the electrons. In this notation we have then

$$\frac{\sigma}{\sigma_0} = 1 - \frac{P}{S} \times \text{constant} = 1 - C\lambda P/S, \quad (2.1)$$

where C will be a dimensionless constant independent of the shape of the specimen.

The rigorous theory of the electrical conductivity of thin metallic films has been given by Fuchs (1938). When the specimen is thick, his expression reduces to

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{8}(1 - \epsilon)\frac{\lambda}{t}, \quad (2.2)$$

where t is the thickness of the film and ϵ the probability of an electron being scattered elastically at the surface. Since the film has two sides, $P/S = 2/t$, and comparing (2.1) and (2.2), we obtain $C = 3(1 - \epsilon)/16$. Hence for a wire of *arbitrary* cross-section,

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{16}(1 - \epsilon)\frac{\lambda P}{S}. \quad (2.3)$$

For a wire of circular cross-section and radius a , $P/S = 2/a$, and thus

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3(1-\epsilon)}{4k},$$

where $k = 2a/\lambda$. This result will be confirmed by the detailed theory given later in this paper.

3. APPROXIMATE METHOD FOR THIN FILMS AND FOR WIRES OF SMALL DIAMETER

When the thickness of the specimen is much less than the mean free path, most paths will start at the surface. Lovell (1936) suggested the following simplifying assumptions:

- (1) All paths start at the surface.
- (2) The effective mean free path may be taken as the average of all free paths, the free path being taken as the distance to the next intersection with the surface or the ordinary mean free path, whichever of these is least.
- (3) The scattering at the surface is inelastic, so that the density of electrons travelling inwards from the surface is uniform.

Using these assumptions, Lovell (1936) gave a theory of the conductivity of a very thin film, and Andrew (1949) one of a very thin wire. These theories may be improved by a more accurate evaluation of the effective mean free path. The probability that an electron will travel a distance r and then suffer a collision between r and $r + dr$ is $e^{-r/\lambda} dr/\lambda$. If the surface cuts short the free paths in such a way that their length is restricted to a maximum value r_0 , say, the effective mean free path (for a fixed value of r_0) of inelastically scattered electrons will be given by

$$\lambda'(r_0) = \int_0^{r_0} r e^{-r/\lambda} dr/\lambda + \int_{r_0}^{\infty} r_0 e^{-r/\lambda} dr/\lambda = \lambda(1 - e^{-r_0/\lambda}). \quad (3.1)$$

In general, r_0 depends on direction, and the effective mean free path is given by the average of (3.1) over the directional co-ordinates. Suppose, for instance, we consider a point on the surface of a thin film. The fraction of paths radiating from it with directions lying between θ and $\theta + d\theta$ and ϕ and $\phi + d\phi$ is $\sin \theta d\theta d\phi/2\pi$, and the distance along any path to its intersection with the opposite boundary is

$$r_0 = t/|\cos \theta|,$$

where t is the thickness of the film. Then

$$\frac{\sigma}{\sigma_0} = \frac{\lambda'}{\lambda} = \int_0^\pi d\phi \int_0^\pi d\theta (1 - e^{-t/\lambda |\cos \theta|}) \sin \theta / 2\pi = 1 - (e^{-k} + k \text{Ei}(-k)), \quad (3.2)$$

where $k = t/\lambda$ and $\text{Ei}(-k) = \int_{\infty}^k e^{-x}/x dx$.

For a wire of diameter d , $r_0 = d \sin \phi / \sin \theta$, and hence

$$\frac{\sigma}{\sigma_0} = \frac{\lambda'}{\lambda} = \int_0^\pi d\phi \int_0^\pi d\theta (1 - e^{-k \sin \phi / \sin \theta}) \sin \theta / 2\pi = 1 - \frac{2}{\pi} \int_0^{\frac{1}{2}\pi} d\phi \int_0^{\frac{1}{2}\pi} d\theta e^{-k \sin \phi / \sin \theta} \sin \theta. \quad (3.3)$$

The integral over θ is of the type T_2 discussed in appendix I. Integrating over ϕ , noting that $\int_0^{\frac{1}{2}\pi} \sin^2 \phi \ln(\sin \phi) d\phi = \frac{1}{8}\pi(1 - 2 \ln 2)$, we obtain for small k ,

$$\frac{\sigma}{\sigma_0} = k - \frac{1}{4}k^2(\ln 1/k - C + 2 \ln 2 + 1). \quad (3.4)$$

(3.2) and (3.4) are similar to the expressions for small k given by the rigorous theory.

PART II. THE RIGOROUS THEORY FOR A WIRE OF CIRCULAR CROSS-SECTION WHEN THE SCATTERING IS INELASTIC

4. THE BOLTZMANN EQUATION

Let $N(\mathbf{r}, \mathbf{v})$ denote the number of electrons per unit volume of ordinary space and per unit volume of velocity space. This distribution function will, in general, be a function of the position, given by the vector $\mathbf{r}$, and of the velocity, given by the vector $\mathbf{v}$. If $N_0(\mathbf{r}, \mathbf{v})$ is the corresponding distribution function when there is no electric field applied, we may write the Boltzmann equation in vector notation as

$$\mathbf{v} \frac{\partial N(\mathbf{r}, \mathbf{v})}{\partial \mathbf{r}} + \dot{\mathbf{v}} \frac{\partial N(\mathbf{r}, \mathbf{v})}{\partial \mathbf{v}} = - \frac{N(\mathbf{r}, \mathbf{v}) - N_0(\mathbf{r}, \mathbf{v})}{\tau}, \quad (4.1)$$

where τ is the time of relaxation.*

We now introduce rectangular Cartesian co-ordinates, both for space and for velocity. The components of the spatial vector $\mathbf{r}$ we take to be x , y and z , with z along the axis of the wire, and correspondingly those of the velocity vector $\mathbf{v}$ to be v_x , v_y and v_z . If F is the electric field applied along the wire, we have $\dot{v}_x = \dot{v}_y = 0$ and $\dot{v}_z = eF/m$, e being the modulus of the electronic charge, and m the mass of the electron.

When stationary conditions have been attained, the distribution function $N = N(x, y, z; v_x, v_y, v_z)$ will not vary along the wire, so that (4.1) reduces to

$$v_x \frac{\partial N}{\partial x} + v_y \frac{\partial N}{\partial y} - \frac{eF}{m} \frac{\partial N_0}{\partial v_z} = - \frac{N - N_0}{\tau}. \quad (4.2)$$

5. THE SOLUTION OF THE BOLTZMANN EQUATION FOR ZERO FIELD

In (4.2), we put $F = 0$ and $N = N_0$. Then

$$v_x \frac{\partial N_0}{\partial x} + v_y \frac{\partial N_0}{\partial y} = 0. \quad (5.1)$$

The subsidiary equations for solving this linear partial differential equation of the first order are

$$dx/v_x = dy/v_y = dN_0/0, \quad (5.2)$$

of which two independent solutions are

$$x/v_x - y/v_y = \text{constant} \quad \text{and} \quad N_0 = \text{constant}. \quad (5.3)$$

* The introduction of such a relaxation time is permissible only if the resistance is largely 'residual', or if the temperature is greater than the Debye characteristic temperature.

Hence
$$N_0(x, y) = N_0(x/v_x - y/v_y). \quad (5.4)$$

If the specimen is very wide in one direction, it is physically obvious that N_0 will not vary appreciably along that direction. It follows from (5.1) that N_0 will not vary along the direction at right angles. Thus if the specimen is wide in one direction, N_0 becomes entirely independent of position. The only function of the type (5.4) which is in agreement with this expected behaviour is that in which N_0 is *always* completely independent of x and y . N_0 is still a function of the velocities. It is essential first to solve this simple case in which there is no applied field, as the result that N_0 is independent of position will be required several times in the ensuing theory.

6. THE SOLUTION OF THE BOLTZMANN EQUATION FOR NON-ZERO FIELDS

For reasonably small electric fields, $N - N_0$ will be small. Putting

$$N - N_0 = n(x, y; v_x, v_y, v_z),$$

and neglecting the product Fn , (4.2) becomes

$$v_x \frac{\partial n}{\partial x} + v_y \frac{\partial n}{\partial y} + \frac{n}{\tau} = \frac{eF}{m} \frac{\partial N_0}{\partial v_z}. \quad (6.1)$$

Since we shall be considering the case of a wire of circular cross-section, it might appear expedient at this point to transform (6.1) into cylindrical co-ordinates.* Actually, this is not the case, as the required results may be obtained more rapidly by finding the general solution of (6.1) as it stands, and then picking out the solutions which possess the required symmetry.

Since N_0 is independent of x and y (§ 5), we may rewrite (6.1) as

$$v_x \frac{\partial n'}{\partial x} + v_y \frac{\partial n'}{\partial y} + \frac{n'}{\tau} = 0, \quad (6.2)$$

where
$$n' = n - \frac{eF\tau}{m} \frac{\partial N_0}{\partial v_z}. \quad (6.3)$$

The subsidiary equations for solving (6.2) are

$$dx/v_x = dy/v_y = -\tau dn'/n', \quad (6.4)$$

of which two independent solutions are

$$x/v_x - y/v_y = \text{constant} \quad \text{and} \quad n' e^{x/\tau v_x} = \text{constant}. \quad (6.5)$$

Thus the general solution of (6.2) is

$$n' = e^{-x/\tau v_x} \phi(x/v_x - y/v_y) = \exp \left\{ -\frac{1}{\tau} \left[\frac{x}{v_x} + \psi \left(\frac{x}{v_x} - \frac{y}{v_y} \right) \right] \right\}, \quad (6.6)$$

where ϕ and ψ are arbitrary functions which may include the velocities as parameters, i.e. $\psi = \psi(v_x, v_y)$.

* This transformation is carried out in appendix II.

7. THE INTRODUCTION OF CYLINDRICAL POLAR CO-ORDINATES

In this section those solutions which possess the required symmetry will be picked out. Let r be the radius vector as measured from the central axis of the wire, and θ an angular variable. Then we take

$$x = r \cos \theta, \quad y = r \sin \theta, \quad z = z. \quad (7.1)$$

In addition, we set up a velocity co-ordinate system for each spatial point (r, θ, z) such that v_r is the velocity resolved along the radial direction (that of increasing r), v_θ that along the tangential direction (that of increasing θ), and v_z that parallel to the axis of the wire. Then

$$\left. \begin{aligned} v_x &= v_r \cos \theta - v_\theta \sin \theta, \\ v_y &= v_r \sin \theta + v_\theta \cos \theta, \\ v_z &= v_z. \end{aligned} \right\} \quad (7.2)$$

Because of the symmetry of a wire of circular cross-section, N must be independent of the angular variable θ , since the radial and tangential velocities v_r and v_θ are so defined that their meaning is not dependent on the value of θ . Thus we must seek solutions of (6.6) which do not contain θ , but which may contain r , v_r and v_θ .

First, we write down those functions of x , y , v_x and v_y which do not contain θ . From (7.1) and (7.2), these are

$$v_x^2 + v_y^2 = v_r^2 + v_\theta^2, \quad (7.3)$$

$$xv_x + yv_y = rv_r, \quad (7.4)$$

$$xv_y - yv_x = rv_\theta. \quad (7.5)$$

One solution of (6.6) which does not contain θ is obtained by putting

$$\psi(\mu) = -\mu v_y^2 / (v_x^2 + v_y^2),$$

giving
$$n' = \exp \left\{ -\frac{1}{\tau} \left(\frac{xv_x + yv_y}{v_x^2 + v_y^2} \right) \right\} = \exp \left(-\frac{1}{\tau} \frac{rv_r}{v_r^2 + v_\theta^2} \right). \quad (7.6)$$

Since arbitrary functions of $v_x^2 + v_y^2 = v_r^2 + v_\theta^2$ and of $xv_y - yv_x = rv_\theta$ may also be introduced into the function ψ of (6.6), the most general solution which is independent of θ is

$$n' = \exp \left\{ -\frac{1}{\tau} \left(\frac{rv_r}{v_r^2 + v_\theta^2} \right) \right\} \chi(rv_\theta, v_r^2 + v_\theta^2), \quad (7.7)$$

where χ is an arbitrary function of the two variables enclosed in the brackets following it. By (6.3), we may take

$$N - N_0 = n = \frac{eF\tau}{m} \frac{\partial N_0}{\partial v_z} \left[1 - f(rv_\theta, v_r^2 + v_\theta^2) \exp \left(-\frac{1}{\tau} \frac{rv_r}{v_r^2 + v_\theta^2} \right) \right], \quad (7.8)$$

where f is a multiple of χ . This solution is confirmed by the work of appendix II. It is worth noting that $f(\alpha, \beta)$ must be an even function of the variable α , since it is physically obvious that n will be an even function of v_θ .

8. THE BOUNDARY CONDITION FOR $r = 0$

In § 9 it will be shown that the boundary condition at the surface of the wire determines the function $f(rv_\theta, v_r^2 + v_\theta^2)$ only for negative radial velocities; for these we write the function as f_- . In this section it will be shown that the function f is the same for positive values of v_r , i.e. $f_+ = f_-$.

For, when an electron crosses the axis $r = 0$, the value of the radial velocity we assign to it changes sign abruptly, being positive when the electron is travelling towards the central axis, and negative when moving away from it. This change in sign has no real physical significance; it is due merely to our convention as to the sign of v_r , and the distribution function suffers no sudden change. Thus

$$N(v_r, v_\theta; r = 0) = N(-v_r, v_\theta; r = 0).$$

Now N_0 and $\partial N_0 / \partial v_z$ are even in v_r ; hence by (7.8)

$$f_+(av_\theta, v_r^2 + v_\theta^2) = f_-(av_\theta, (-v_r)^2 + v_\theta^2).$$

This identity can be satisfied for all v_r and v_θ only if the functions are identical, i.e. $f_+ = f_-$. Thus the same function must be inserted in (7.8) for positive as for negative values of the radial velocity.

9. THE BOUNDARY CONDITION FOR $r = a$

When the scattering at the surface is completely inelastic, the electrons returning from the surface must have directions distributed at random. Thus if a is the radius of the wire, $N(-|v_r|, v_\theta; r = a)$ must possess no directional properties. This condition can be satisfied only if $n(-|v_r|, v_\theta; r = a)$ vanishes, since it contains as a factor the quantity $\partial N_0 / \partial v_z$, which certainly has directional properties. Thus by (7.8),

$$1 - f(av_\theta, v_r^2 + v_\theta^2) \exp\left(\frac{a|v_r|}{\tau(v_r^2 + v_\theta^2)}\right) = 0. \quad (9.1)$$

(9.1) can be satisfied for all v_r and v_θ only by taking

$$f(\alpha, \beta) = \exp\{-(a^2\beta - \alpha^2)^{\frac{1}{2}}/\tau\beta\}, \quad (9.2)$$

the *positive* square root always being understood. By (7.8)

$$n = \frac{eF\tau}{m} \frac{\partial N_0}{\partial v_z} \left[1 - \exp\left\{-\frac{rv_r + \{a^2(v_r^2 + v_\theta^2) - r^2v_\theta^2\}^{\frac{1}{2}}}{\tau(v_r^2 + v_\theta^2)}\right\}\right]. \quad (9.3)$$

10. THE CALCULATION OF THE CURRENT DENSITY AND THE CONDUCTIVITY

At each point in ordinary space, the following co-ordinate system is chosen for the velocities:

$$v_r = v \sin \theta \sin \phi, \quad v_\theta = v \sin \theta \cos \phi, \quad v_z = v \cos \theta, \quad (10.1)$$

with $0 < \theta < \pi$ and $0 < \phi < 2\pi$. The element of 'volume' in velocity space is

$$v^2 \sin \theta dv d\theta d\phi.$$

The current density $j(r)$ at a radial distance r from the central axis of the wire will be

$$j(r) = e \int_{v=0}^{\infty} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} v_z n(v, r) v^2 \sin \theta dv d\theta d\phi. \quad (10.2)$$

Since
$$v_z \frac{\partial N_0}{\partial v_z} = \frac{v_z^2}{v} \frac{\partial N_0}{\partial v} = v \cos^2 \theta \frac{\partial N_0}{\partial v}, \quad (10.3)$$

$$j(r) = \frac{e^2 F \tau}{m} \int_0^{\infty} v^3 \frac{\partial N_0}{\partial v} dv \int_0^{\pi} d\theta \cos^2 \theta \sin \theta \int_0^{2\pi} d\phi \left[1 - \exp \left\{ - \frac{r \sin \phi + (a^2 - r^2 \cos^2 \phi)^{\frac{1}{2}}}{\tau v \sin \theta} \right\} \right]. \quad (10.4)$$

We assume that $N_0(v)$ is isotropic in velocity space, i.e. that the Fermi surface is spherical; then the limits of the three integrations in (10.4) are independent. Introducing the usual assumption that $N_0(v)$ varies appreciably only at the surface of the velocity distribution, we have, on putting $\tau v_{\text{Fermi}} = \lambda$ (the mean free path) and taking j_0 to be the current density which would exist in the bulk metal for the same applied electric field,

$$\frac{j(r)}{j_0} = \frac{3}{4\pi} \int_0^{\pi} d\theta \cos^2 \theta \sin \theta \int_0^{2\pi} d\phi \left[1 - \exp \left\{ - \frac{r \sin \phi + (a^2 - r^2 \cos^2 \phi)^{\frac{1}{2}}}{\lambda \sin \theta} \right\} \right]. \quad (10.5)$$

This may be written in the alternative form

$$\frac{j(r)}{j_0} = 1 - \frac{6}{\pi} \int_0^{\frac{1}{2}\pi} d\theta \cos^2 \theta \sin \theta \int_0^{\frac{1}{2}\pi} d\phi \cosh \left(\frac{r \sin \phi}{\lambda \sin \theta} \right) \exp \left\{ - \frac{(a^2 - r^2 \cos^2 \phi)^{\frac{1}{2}}}{\lambda \sin \theta} \right\}. \quad (10.6)$$

Experimental measurements have been made not of $j(r)/j_0$ but of the apparent overall conductivity of the wire. Since the area of an annulus is proportional to $r dr$, we have for the ratio of the apparent conductivities

$$\frac{\sigma}{\sigma_0} = \int_0^a j(r) r dr / \int_0^a j_0 r dr = \frac{2}{a^2} \int_0^a (j(r)/j_0) r dr. \quad (10.7)$$

One integration of (10.7) may be achieved by putting $x = r \cos \phi$ and $y = r \sin \phi$, and replacing $r dr d\phi$ by $dx dy$. Then

$$\begin{aligned} \frac{\sigma}{\sigma_0} &= 1 - \frac{3}{\pi a^2} \int_0^{\frac{1}{2}\pi} d\theta \cos^2 \theta \sin \theta \int_{-a}^a dx \int_{-(a^2-x^2)^{\frac{1}{2}}}^{(a^2-x^2)^{\frac{1}{2}}} dy \exp \left\{ - \frac{y + (a^2-x^2)^{\frac{1}{2}}}{\lambda \sin \theta} \right\} \\ &= 1 - \frac{12}{\pi k} \int_0^{\frac{1}{2}\pi} d\theta \cos^2 \theta \sin^2 \theta \left(1 - \int_0^{\frac{1}{2}\pi} e^{-k \sin \psi / \sin \theta} \sin \psi d\psi \right), \end{aligned}$$

where $k = 2a/\lambda$ and $x = a \cos \psi$. In the notation of appendix I,

$$\begin{aligned} \frac{\sigma}{\sigma_0} &= 1 - \frac{12}{\pi k} \left[S_5(0) - \int_0^{\frac{1}{2}\pi} \sin \psi S_5(k \sin \psi) d\psi \right] \\ &= 1 - \frac{12}{\pi} \int_0^{\frac{1}{2}\pi} \cos^2 \psi S_4(k \sin \psi) d\psi = 1 - \frac{12}{\pi} \int_0^1 (1-x^2)^{\frac{1}{2}} S_4(kx) dx, \end{aligned} \quad (10.8)$$

on integration by parts.

Exact integrations of (10.6) and (10.8) are impracticable, but approximations may be found which are adequate in the two limiting cases, when $2a/\lambda = k \gg 1$ and when $k \ll 1$. In the intermediary case where $k \sim 1$, it is necessary to resort to numerical integration.

11. VALUES FOR THICK WIRES ($k \gg 1$)

When $2a \gg \lambda$, $\cosh(r \sin \phi / \lambda \sin \theta) \simeq \frac{1}{2} \exp(r \sin \phi / \lambda \sin \theta)$,

and (10.6) shows that $j(r)/j_0 \simeq 1$ unless

$$r \sin \phi - (a^2 - r^2 \cos^2 \phi)^{\frac{1}{2}} \simeq 0 \quad \text{i.e. } r \simeq a.$$

Thus when $k \gg 1$, deviations from the current density which would exist in the bulk metal subjected to the same field occur only at the surface. This confirms the argument of § 2. In (10.6) we take $r = a(1 - z)$ where z is small. Then

$$\frac{j(r)}{j_0} = 1 - \frac{3}{\pi} \int_0^{\frac{1}{2}\pi} d\theta \cos^2 \theta \sin \theta \int_0^{\frac{1}{2}\pi} d\phi \exp \left\{ -\frac{kz}{2 \sin \theta \sin \phi} \right\} \left(1 + \frac{kz^2 \cos^2 \phi}{4 \sin \theta \sin^3 \phi} \right). \quad (11.1)$$

(11.1) shows that for large values of k , the fall in the current density near the surface is roughly exponential. Near the surface, the current density is a function of $kz = k(1 - r/a)$ only. Table 1 shows some values of $j(r)/j_0$ obtained by numerical computation.

Expanding the root in (10.8) and integrating over x , we obtain*

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{4k} + \frac{3}{8k^3} \dots \quad (11.2)$$

This result agrees with that of § 2 for $k \gg 1$ and $\epsilon = 0$.

TABLE 1. $k \gg 1$

$k(1 - r/a)$	$j(r)/j_0$
0	0.500
0.2	0.671
0.5	0.759
1	0.887
2	0.9516
5	0.99523
10	0.999842
∞	1.00000

12. VALUES FOR THIN WIRES ($k \ll 1$)

When $2a \gg \lambda$, we may expand the exponential and the hyperbolic cosine appearing in (10.6), and obtain

$$\frac{j(r)}{j_0} \simeq \frac{6}{\pi \lambda} \int_0^{\frac{1}{2}\pi} d\theta \cos^2 \theta \int_0^{\frac{1}{2}\pi} (a^2 - r^2 \cos^2 \phi)^{\frac{1}{2}} d\phi = \frac{3k}{4} E(r/a), \quad (12.1)$$

* The first two terms, and the absence of a term in k^{-2} , may also be obtained by integrating (11.1) over z .

where E is a complete elliptic integral, with limiting values $E(0) = \frac{1}{2}\pi$ and $E(1) = 1$. (12.1) shows that the actual *distribution* of the current density across the wire becomes independent of k when $k \ll 1$. Table 2 gives the values of $j(r)/kj_0$ obtained from (12.1).

TABLE 2. $k \ll 1$

r/a	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$j(r)/kj_0$	1.178	1.159	1.102	1.038	0.750

More accurate values may easily be obtained for the current density at the centre of the wire and at the surface, as the integrals involved are of the form S_4 discussed in appendix I. For when $r = 0$ (10.6) reduces to

$$\begin{aligned} j(0)/j_0 &= 1 - 3 \int_0^{\frac{1}{2}\pi} e^{-\frac{1}{2}k \sin \theta} \cos^2 \theta \sin \theta d\theta = 1 - 3S_4(\tfrac{1}{2}k) \\ &\simeq \frac{3\pi k}{4} - \frac{3k^2}{8} (\ln 1/k + 2 \ln 2 - C + \tfrac{1}{2}) - \frac{\pi k^3}{32}. \end{aligned} \tag{12.2}$$

When $r = a$, (10.6) reduces to

$$\begin{aligned} j(a)/j_0 &= 1 - \frac{3}{\pi} \int_0^{\frac{1}{2}\pi} d\phi \int_0^{\frac{1}{2}\pi} d\theta \cos^2 \theta \sin \theta e^{-k \sin \phi / \sin \theta} = 1 - \frac{3}{\pi} \int_0^{\frac{1}{2}\pi} d\phi S_4(k \sin \phi) \\ &\simeq \frac{3k}{4} - \frac{3k^2}{8} (\ln 1/k + 2 \ln 2 - C) - \frac{k^3}{6}. \end{aligned} \tag{12.3}$$

Since by (10.6) $j(r)/j_0$ must be even in r , (12.2) and (12.3) suggest the interpolation formula

$$\frac{j(r)}{j_0} \simeq \tfrac{3}{4}kE(r/a) - \tfrac{3}{8}k^2[\ln 1/k + 2 \ln 2 - C + \tfrac{1}{2}(1 - r^2/a^2)]. \tag{12.4}$$

By integration of (12.4), or more rigorously by inserting the approximate values of $S_4(k \sin \psi)$ in (10.8), we obtain

$$\frac{\sigma}{\sigma_0} \simeq k - \frac{3k^2}{8} (\ln 1/k + 2 \ln 2 - C + \tfrac{1}{4}) - \frac{2k^3}{15} = k - \frac{3k^2}{8} (\ln 1/k + 1.059) - \frac{2k^3}{15}. \tag{12.5}$$

13. TABLES OF VALUES OF $j(r)/j_0$ AND σ/σ_0

For $k \sim 1$ the approximate expressions of §§ 11 and 12 are inadequate, and it becomes necessary to resort to numerical integration. The ratio $j(r)/j_0$ given by (10.6) was accordingly evaluated with $r/a = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ and 1, the calculations being carried out in each case for $k = \frac{1}{2}, 1$ and 2. The values of σ/σ_0 determined from these values of the current density by using Gregory's integration formula agree closely with those obtained by numerical integration of (10.8) for the same values of k , and given in table 4.

TABLE 3. $k \sim 1$

k	$j(r)/j_0$					σ/σ_0 (from $j(r)/j_0$)
	$r/a = 0$	$r/a = \frac{1}{4}$	$r/a = \frac{1}{2}$	$r/a = \frac{3}{4}$	$r/a = 1$	
$\frac{1}{2}$	0.386	0.371	0.356	0.311	0.209	0.312
1	0.599	0.574	0.554	0.484	0.307	0.482
2	0.813	0.797	0.761	0.671	0.387	0.661

TABLE 4. k = DIAMETER/MEAN FREE PATH FOR A WIRE,
AND THICKNESS/MEAN FREE PATH FOR A FILM

k	σ/σ_0 (wire)		σ/σ_0 (film)
	$\epsilon = 0$	$\epsilon = \frac{1}{2}$	$\epsilon = 0$
0.001	0.00100	0.00297	0.00550
0.002	0.00199	0.00589	0.00996
0.005	0.00494	0.01442	0.02145
0.01	0.0098	0.0280	0.0378
0.02	0.0192	0.0534	0.0653
0.05	0.0464	0.1203	0.1300
0.1	0.0873	0.2050	0.212
0.2	0.158	0.331	0.333
0.5	0.318	0.544	0.526
1	0.489	0.703	0.684
2	0.678	0.828	0.819
5	0.853	0.926	0.925
10	0.925	0.963	0.9625
20	0.9625	0.98125	0.98125
50	0.9850	0.99250	0.99250
100	0.9925	0.99625	0.99625

PART III. THE ELECTRICAL CONDUCTIVITY OF A THIN WIRE WHEN THE SCATTERING IS NOT INELASTIC

14. THE BOUNDARY CONDITION FOR $r = a$, WHEN THE SCATTERING IS PERFECTLY ELASTIC

If we suppose that the scattering of the electrons at the surface of the metal is always specular, we may easily show that the conductivity of the wire is the same as that of the bulk metal. For elastic scattering, all electrons incident on the surface with radial velocity v_r will be reflected with radial velocity $-v_r$. Thus

$$N(v_r, v_\theta; r=a) = N(-v_r, v_\theta; r=a). \tag{14.1}$$

Since N_0 and $\partial N_0/\partial v_z$ are even in v_r , it follows from (7.8) that

$$f(av_\theta, v_r^2 + v_\theta^2) \exp\left\{\frac{-av_r}{\tau(v_r^2 + v_\theta^2)}\right\} = f(av_\theta, v_r^2 + v_\theta^2) \exp\left\{\frac{av_r}{\tau(v_r^2 + v_\theta^2)}\right\}, \tag{14.2}$$

which can be satisfied for all values of v_r and v_θ only if $f \equiv 0$, giving $n = (eF\tau/m) \partial N_0/\partial v_z$, which, being independent of the radius of the wire, is the distribution function for the bulk metal.

15. THE BOUNDARY CONDITION FOR $r = a$, WHEN THE SCATTERING IS PARTIALLY ELASTIC

In this section it is assumed that the probability of elastic scattering at the surface is ϵ , and that of inelastic scattering is $(1 - \epsilon)$. When $\epsilon = 0$, the results obtained will reduce to those of the second part of this paper, and when $\epsilon = 1$ to those of § 14.

The number of electrons at the surface which are travelling inwards with speed v_r will be the sum of the number of electrons which had previously been travelling

outwards with speed v_r and which have been specularly reflected, and the number which have been scattered inelastically at the surface. Thus, if the distribution function for the inelastically scattered electrons is represented by g ,

$$N(-v_r, v_\theta; r=a) = \epsilon N(v_r, v_\theta; r=a) + g,$$

whence

$$g = (1-\epsilon)N_0 + n(-v_r, v_\theta; r=a) - \epsilon n(v_r, v_\theta; r=a) \quad (15.1)$$

$$= (1-\epsilon)N_0 + \frac{eF\tau}{m} \frac{\partial N_0}{\partial v_z} \left[1 - \epsilon - f(av_\theta, v_r^2 + v_\theta^2) \left\{ \exp\left(\frac{av_r}{\tau(v_r^2 + v_\theta^2)}\right) - \epsilon \exp\left(\frac{-av_r}{\tau(v_r^2 + v_\theta^2)}\right) \right\} \right] \quad (15.2)$$

by (7.8). Since $\partial N_0/\partial v_z$ depends on direction, it is only possible that g shall be independent of direction—as postulated—if the quantity in square brackets vanishes. This condition can be satisfied for all v_r and v_θ only by taking

$$(1-\epsilon)f^{-1}(\alpha, \beta) = \exp\left\{\frac{(a^2\beta - \alpha^2)^{\frac{1}{2}}}{\tau\beta}\right\} - \epsilon \exp\left\{-\frac{(a^2\beta - \alpha^2)^{\frac{1}{2}}}{\tau\beta}\right\}. \quad (15.3)$$

Hence, by (7.8),

$$n = \frac{eF\tau}{m} \frac{\partial N_0}{\partial v_z} \left[1 - \frac{(1-\epsilon) \exp[-\{rv_r + (a^2(v_r^2 + v_\theta^2) - r^2v_\theta^2)^{\frac{1}{2}}\}]/\tau(v_r^2 + v_\theta^2)]}{1 - \epsilon \exp[-2\{a^2(v_r^2 + v_\theta^2) - r^2v_\theta^2\}^{\frac{1}{2}}]/\tau(v_r^2 + v_\theta^2)} \right]. \quad (15.4)$$

On introducing the velocity co-ordinate system (10.1), and formally expanding the denominator, this reduces to

$$n = \frac{eF\tau}{m} \frac{\partial N_0}{\partial v_z} \left[1 - (1-\epsilon) \sum_{v=0}^{\infty} \epsilon^v \exp\left\{-\frac{r \sin \phi + (2v+1)(a^2 - r^2 \cos^2 \phi)^{\frac{1}{2}}}{\tau v \sin \theta}\right\} \right]. \quad (15.5)$$

16. THE CALCULATION OF THE CURRENT DENSITY AND THE CONDUCTIVITY

By the methods of § 10, we obtain

$$\begin{aligned} \frac{j(r)}{j_0} &= 1 - \frac{6}{\pi} (1-\epsilon) \sum_{v=0}^{\infty} \epsilon^v \int_0^{\frac{1}{2}\pi} d\theta \cos^2 \theta \sin \theta \\ &\quad \times \int_0^{\frac{1}{2}\pi} d\phi \cosh\left(\frac{r \sin \phi}{\lambda \sin \theta}\right) \exp\left\{-\frac{(2v+1)(a^2 - r^2 \cos^2 \phi)^{\frac{1}{2}}}{\lambda \sin \theta}\right\} \end{aligned} \quad (16.1)$$

$$\text{and} \quad \frac{\sigma}{\sigma_0} = 1 - \frac{12}{\pi} (1-\epsilon)^2 \sum_{v=1}^{\infty} v \epsilon^{v-1} \int_0^1 (1-x^2)^{\frac{1}{2}} S_4(vkx) dx. \quad (16.2)$$

$$\text{Thus} \quad \frac{\sigma}{\sigma_0} = (1-\epsilon)^2 \sum_{v=1}^{\infty} v \epsilon^{v-1} [\sigma(vk)/\sigma_0]_{\epsilon=0}. \quad (16.3)$$

This relation (16.3) is also valid for the case of a thin film.

When $k \gg 1$, a similar argument to that of § 11 shows that the deviation of the current density from that of the bulk metal is significant only when $r \simeq a$ and $v = 0$ in (16.1). $j(r)/j_0$ may therefore be obtained from table 1 by using the relation

$$\frac{j(r)}{j_0} = 1 - (1-\epsilon) \left[1 - \left(\frac{j(r)}{j_0} \right)_{\epsilon=0} \right]. \quad (16.4)$$

By expanding the root in (16.2), we obtain*

$$\frac{\sigma}{\sigma_0} \simeq 1 - \frac{3}{4k}(1-\epsilon) + \frac{3}{8k^3}(1-\epsilon)^2 \sum_1^{\infty} \epsilon^{v-1}/v^2. \quad (16.5)$$

For $k \gg 1$, this agrees with the work of § 2.

When $k \ll 1$, similar arguments to those of § 12 lead to the expressions

$$\frac{j(r)}{j_0} \sim \frac{1+\epsilon}{1-\epsilon} \frac{3k}{4} E(r/a), \quad (16.6)$$

$$\text{and} \quad \frac{\sigma}{\sigma_0} \simeq \frac{1+\epsilon}{1-\epsilon} k - \frac{3k^2}{8} \left[\frac{1+4\epsilon+\epsilon^2}{(1-\epsilon)^2} (\ln 1/k + 1.059) - (1-\epsilon)^2 \sum_1^{\infty} v^3 \epsilon^{v-1} \ln v \right] - \frac{2k^3(1+11\epsilon+11\epsilon^2+\epsilon^3)}{15(1-\epsilon)^3}. \quad (16.7)$$

For $\epsilon = \frac{1}{2}$, (16.5) and (16.7) reduce to

$$\frac{\sigma}{\sigma_0} \simeq 1 - \frac{0.375}{k} + \frac{0.109}{k^3} \quad (k \gg 1), \quad (16.8)$$

$$\text{and} \quad \frac{\sigma}{\sigma_0} \simeq 3k - 11.22k^2(\log 1/k - 0.245) - 9.74k^3 \quad (k \ll 1). \quad (16.9)$$

The values given in table 4 for $\epsilon = \frac{1}{2}$, $2 \geq k \geq 0.1$, were calculated directly from the summation (16.3).

PART IV. COMPARISON WITH EXPERIMENT

The recent experimental results obtained by Andrew (1949) for a thin mercury wire were compared with our theory by plotting the experimental values of $\log(\sigma_{\text{room}}/\sigma)$ against $\log(\text{diameter})$ for each temperature, and fitting to the curves obtained the theoretical curve of $\log(\sigma_0/\sigma)$ plotted against $\log k$. These curves, and the scatter of the experimental points about them, are very similar to those given in Andrew's paper, and need not be reproduced here. From the fitting, values of λ (the mean free path) and σ_0 (the bulk conductivity) may be derived for each temperature. Since the bulk conductivity is proportional to the mean free path, $\lambda(T)/\sigma_0(T)$ should be independent of temperature. The average values of the observations on *all* the wires gave the results

$$\begin{aligned} \text{Mean free path} \times \text{resistivity of mercury} &= \lambda/\sigma_0 \sim 1.7 \times 10^{-11} \text{ ohm cm.}^2 \quad \text{if } \epsilon = 0, \\ &\text{and } 5.0 \times 10^{-11} \text{ ohm cm.}^2 \quad \text{if } \epsilon = \frac{1}{2}. \end{aligned}$$

The experimental values are not sufficiently accurate to give much indication of the value of ϵ . The values of λ/σ_0 for $\epsilon \neq 0$ fall with rise in temperature, suggesting that ϵ increases with temperature. This is not plausible on theoretical grounds, since the diffuseness of the reflexion at the surface is probably due to the following causes:

(1) The thermal vibration of the surface atoms. It is known that at low temperatures the maximum angle of scatter at a collision in the bulk metal is about T/θ ,

* The first two terms may also be obtained from (16.4) and (11.2).

where θ is the Debye characteristic temperature. A similar relation will probably hold for the surface atoms, so that the scattering will become more diffuse the higher the temperature.

(2) The number of mobile surface defects of various kinds. This number will also increase with temperature, giving more diffuse scattering.

(3) The disturbance in the electric potential due to surface strains and impurities.

At the very low temperatures of Andrew's experiments, the temperature variation in ϵ due to these causes would probably be small, but if it were appreciable it would be such that ϵ fell with rise in temperature. Since the fitting for $\epsilon \neq 0$ suggests the opposite temperature variation for ϵ , the value $\epsilon = 0$ is probably the more likely.

In obtaining the results given above we have taken into account the observations on *all* the wires used by Andrew. However, there are two reasons for doubting the reliability of the experimental results for the wires of diameter less than 2×10^{-3} cm.:

(1) These specimens were prepared in a different way from the remainder.

(2) The values of σ/σ_0 calculated in the present paper do not differ appreciably* from those obtained from the Nordheim (1934) relation, which, written in the form

$$\frac{1}{\sigma} = \frac{1}{\sigma_0} + \frac{\alpha(\lambda/\sigma_0)}{2a},$$

is similar to Matthiessen's rule, with $\alpha(\lambda/\sigma_0)/2a$ playing the role of a residual resistance independent of temperature.

Since the resistivities of the thinnest wires used by Andrew do not in fact vary with temperature in the manner predicted by Matthiessen's rule, they cannot be consistent with the above theoretical prediction. If these very thin wires are not taken into consideration, the results obtained are $\lambda/\sigma_0 \sim 3.6 \times 10^{-11}$ ohm cm.² if $\epsilon = 0$, and 8.7×10^{-11} ohm cm.² if $\epsilon = \frac{1}{2}$.

It must be concluded that the value of λ/σ_0 derived from Andrew's experiments depends markedly both on the value of ϵ assumed, and on the trust placed in the results for the very thin wires. It is, however, satisfactory that the values agree in order of magnitude with those obtained by Reuter & Sondheimer (1948) from Pippard's (1947) measurements of the anomalous skin effect in mercury. Fitting the experimental points to lie between the theoretical curves for $\epsilon = 0$ and $\epsilon = 1$, they obtain the result $\lambda/\sigma_0 \sim 2.7 \times 10^{-11}$ ohm cm.².

Finally, it should be mentioned that mercury probably does not have a spherical Fermi surface, and experimental results for a more suitable metal would be required for a wholly satisfactory comparison with the theory.

APPENDIX I. THE EVALUATION OF INTEGRALS OF THE TYPE S_n AND T_n

We define

$$S_n = \int_0^{\frac{1}{2}\pi} e^{-u/\sin\theta} \cos^2\theta \sin^{n-3}\theta d\theta \quad \text{and} \quad T_n = \int_0^{\frac{1}{2}\pi} e^{-u/\sin\theta} \sin^{n-1}\theta d\theta.$$

By differentiation with respect to the parameter u ,

$$S_{n-1} = -\partial S_n / \partial u \quad \text{and} \quad T_{n-1} = -\partial T_n / \partial u. \quad \text{I (1)}$$

* With $\alpha = 1$ and $\epsilon = 0$, the difference is always less than about 5%.

Substituting $x = \operatorname{cosec} \theta$, we have

$$S_n = \int_1^\infty e^{-ux} (x^2 - 1)^{\frac{1}{2}} x^{-n} dx \quad \text{and} \quad T_n = \int_1^\infty e^{-ux} (x^2 - 1)^{-\frac{1}{2}} x^{-n} dx. \quad \text{I (2)}$$

By I (2),

$$S_n = T_{n-2} - T_n. \quad \text{I (3)}$$

But by integrating S_n by parts

$$S_n = \frac{1}{u} (nT_{n+1} - (n-1)T_{n-1}). \quad \text{I (4)}$$

On eliminating S_n between I (3) and I (4), we obtain a recurrence relation

$$-uT_{n-3} - (n-2)T_{n-2} + uT_{n-1} + (n-1)T_n = 0. \quad \text{I (5)}$$

By I (1), this may be transformed into the differential equation

$$u \frac{\partial^3 T_n}{\partial u^3} - (n-2) \frac{\partial^2 T_n}{\partial u^2} - u \frac{\partial T_n}{\partial u} + (n-1)T_n = 0, \quad \text{I (6)}$$

which reduces to the following simpler result when $n = 1$:

$$u \frac{\partial^2 T_0}{\partial u^2} + \frac{\partial T_0}{\partial u} - uT_0 = 0. \quad \text{I (7)}$$

This equation has a solution $T_0 = AK_0(u)$, where K_0 is a Bessel function of imaginary argument, and A a constant. The Bessel function I_0 , which also satisfies I (7), is inadmissible, as the integral $T_0 \rightarrow 0$ as $u \rightarrow \infty$, whilst $I_0 \rightarrow \infty$. $K_0(u)$ is related to the Hankel function of the first kind by the relation $K_0(u) = \frac{1}{2}i\pi H_1(iu)$.

The constant A is found by evaluating T_0 for very large values of its argument, values sufficiently large for the integrand to be appreciable only for $x \sim 1$, so that we may make the substitution $x = 1 + y$, giving

$$T_0(u) = \int_1^\infty \frac{e^{-ux} dx}{(x^2 - 1)^{\frac{1}{2}}} \simeq \frac{e^{-u}}{\sqrt{2}} \int_0^\infty y^{-\frac{1}{2}} e^{-uy} dy = e^{-u} (\pi/2u)^{\frac{1}{2}}. \quad \text{I (8)}$$

Since this is the same as the leading term in the asymptotic expansion for $K_0(u)$ when $u \gg 1$ (cf. Schelkunoff 1943), $A = 1$, and hence $T_0 \equiv K_0$.

Next, we evaluate $T_n(0) = \int_0^{\frac{1}{2}\pi} \sin^{n-1} \theta d\theta$, obtaining

$$T_0(0) = \infty, \quad T_1(0) = \frac{1}{2}\pi, \quad T_2(0) = 1, \quad T_3(0) = \frac{1}{4}\pi \quad \text{and} \quad T_4(0) = \frac{2}{3}. \quad \text{I (9)}$$

For small u (Schelkunoff 1943)

$$-T_0 = -K_0 \simeq \ln u + C - \ln 2. \quad \text{I (10)}$$

Integrating successively, and finding the integration constants by putting $u = 0$ and comparing with I (9), we obtain

$$-T_2 = \frac{1}{2}u^2 \ln u + \frac{1}{2}u^2(C - \ln 2 - \frac{3}{2}) + \frac{1}{2}\pi u - 1 \quad \text{I (11)}$$

and

$$-T_4 = \frac{u^4 \ln u}{24} + \frac{u^4}{24}(C - \ln 2 - \frac{25}{12}) + \frac{\pi}{12}u^3 - \frac{u^2}{2} + \frac{\pi u}{4} - \frac{2}{3}. \quad \text{I (12)}$$

By I (3)

$$S_4 = T_2 - T_4 \simeq \frac{1}{3} - \frac{1}{2}u^2 \ln u - \frac{1}{4}\pi u - \frac{1}{2}u^2(C - \ln 2 - \frac{1}{2}) + \frac{\pi u^3}{12}. \quad \text{I (13)}$$

The accuracy of this expression is sufficient to justify the inclusion of the terms in k^3 in (12.5) and (16.7) for $k \leq 0.1$.

APPENDIX II. THE TRANSFORMATION OF (6.1) INTO CYLINDRICAL POLAR CO-ORDINATES

From (7.1) and (7.2),

$$\frac{\partial r}{\partial x} = \cos \theta, \quad \frac{\partial \theta}{\partial x} = -\frac{\sin \theta}{r}, \quad \frac{\partial v_r}{\partial \theta} = v_\theta, \quad \frac{\partial v_\theta}{\partial \theta} = -v_r, \quad \frac{\partial v_r}{\partial r} = \frac{\partial v_\theta}{\partial r} = 0, \quad \text{II (1)}$$

whence
$$\frac{\partial v_r}{\partial x} = \frac{\partial r}{\partial x} \frac{\partial v_r}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial v_r}{\partial \theta} = -\frac{\sin \theta}{r} v_\theta \quad \text{II (2)}$$

and
$$\frac{\partial v_\theta}{\partial x} = \frac{\partial r}{\partial x} \frac{\partial v_\theta}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial v_\theta}{\partial \theta} = \frac{\sin \theta}{r} v_r. \quad \text{II (3)}$$

Thus
$$\begin{aligned} \frac{\partial}{\partial x} &\equiv \frac{\partial r}{\partial x} \frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x} \frac{\partial}{\partial \theta} + \frac{\partial v_r}{\partial x} \frac{\partial}{\partial v_r} + \frac{\partial v_\theta}{\partial x} \frac{\partial}{\partial v_\theta} \\ &\equiv \cos \theta \frac{\partial}{\partial r} - \frac{\sin \theta}{r} \left(\frac{\partial}{\partial \theta} + v_\theta \frac{\partial}{\partial v_r} - v_r \frac{\partial}{\partial v_\theta} \right). \end{aligned} \quad \text{II (4)}$$

Similarly
$$\frac{\partial}{\partial y} \equiv \sin \theta \frac{\partial}{\partial r} + \frac{\cos \theta}{r} \left(\frac{\partial}{\partial \theta} + v_\theta \frac{\partial}{\partial v_r} - v_r \frac{\partial}{\partial v_\theta} \right). \quad \text{II (5)}$$

Hence (6.1) transforms to

$$v_r \frac{\partial n}{\partial r} + \frac{v_\theta^2}{r} \frac{\partial n}{\partial v_r} - \frac{v_\theta v_r}{r} \frac{\partial n}{\partial v_\theta} + \frac{n}{\tau} = \frac{eF}{m} \frac{\partial N_0}{\partial v_z}, \quad \text{II (6)}$$

where $n = n(r, \theta; v_r, v_\theta, v_z)$, the term $(v_\theta/r) (\partial n / \partial \theta)$ vanishing owing to the symmetry of the wire. It may easily be verified that (7.8) satisfies this differential equation.

The method used in this paper is essentially the same as that given by Fuchs in his discussion of the electrical conductivity of a thin film, and the present paper owes much to his work. The present author also wishes to record his acknowledgement of the discussions he has had with members of this laboratory, particularly Mr R. G. Chambers—who achieved the partial integration of (10.7), and was thus able to compute the accurate values of σ/σ_0 given in table 4, and who also contributed to the discussion of Andrew's experimental results—Mr A. B. Pippard, Mr A. R. Fraser and Dr E. H. Sondheimer.

REFERENCES

- Andrew, E. R. 1949 *Proc. Phys. Soc.* **62**, 77.
 Fuchs, K. 1938 *Proc. Camb. Phil. Soc.* **34**, 100.
 Lovell, A. C. B. 1936 *Proc. Roy. Soc. A*, **157**, 311.
 Nordheim, L. 1934 *Actes Scientifiques et Industrielles*, no. 131, Paris. (Cf. also Eucken, A. & Förster, F., 1934, *Nachr. Ges. Wiss. Göttingen*, **1**, 43, 129; Nordheim, L., 1931, *Ann. Phys., Lpz.*, **9**, 607; Tisza, L., 1931, *Naturwissenschaften*, **19**, 86.)
 Pippard, A. B. 1947 *Proc. Roy. Soc. A*, **191**, 385.
 Reuter, G. E. H. & Sondheimer, E. H. 1948 *Proc. Roy. Soc. A*, **195**, 336.
 Schelkunoff, S. A. 1943 *Electromagnetic waves*, p. 51. New York: Van Nostrand.

Spectroscopic studies of low-pressure flames

III. Effective rotational temperatures and excitation mechanism for C₂ bands

BY A. G. GAYDON AND H. G. WOLFHARD

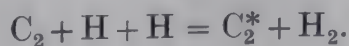
Chemical Engineering Department, Imperial College, London

*(Communicated by Sir Alfred Egerton, F.R.S.—Received 29 October 1949
—Read 2 February 1950)*

Intensity measurements on the Swan bands of C₂ give effective rotational temperatures for a number of flames. For oxygen/acetylene at 1 atm. the value is 4950° K, falling to 3800° K at 2 mm. For acetylene/air it is 3400° K at 1 atm., falling a little at low pressure. In discussing the excitation process, reasons are given against its being a true chemiluminescence. It is believed that the C₂ radicals are formed with high rotational energy and are electronically excited by collision with energy-rich molecules present in the flame and responsible for the high excitation temperatures previously recorded. These energy-rich molecules probably possess excess vibrational energy.

INTRODUCTION

To make the best use of flame spectra for studying the fundamentals of combustion processes it is necessary to find out as much as possible about the mode of excitation of the various radicals such as OH, CH and C₂ whose emission spectra are observed. Only if we are sure that excitation is due solely to normal thermal causes can intensities be used as a criterion of the concentration of the radicals in the flame. If there is direct chemiluminescence, i.e. if the radicals are formed in the electronically excited condition as the result of the chemical action in which they are produced, then the occurrence of this chemiluminescence and its strength may give interesting information about processes taking place during the combustion, but cannot give any value for the concentration of the normal unexcited radicals which may be present. Another possibility is that radicals may be formed in the normal unexcited state but be subsequently excited by chemiluminescence by acting as third body in some other process, such, for example, as



A further possibility is that the radicals may be excited to an abnormal extent by collision with other activated atoms or molecules formed by chemical processes of the general type



In this case the strength of emission would be as much a measure of the concentration of the active species A* as of the radicals C₂.

The Swan bands of C₂ are the most prominent feature of the visible spectra of the flames of hydrocarbons and many organic compounds. They are of especial interest because it has been suggested that their occurrence may be associated with the formation of solid carbon particles. The control of carbon formation in flames is of

course of major importance for many industrial applications, and any information about it would be of considerable value. It is as yet not certain whether solid carbon may be formed by polymerization of C_2 radicals, as suggested by Smith (1940), whether the C_2 is formed by vaporization of carbon particles or incipient carbon particles formed by polymerization and cracking of the hydrocarbons, or whether the solid carbon particles and C_2 radicals are formed independently by different and perhaps competing mechanisms. The experimental work described in this and the following part was undertaken with the object of obtaining as many factual data as possible for attacking this problem.

Measurement of the effective rotational temperature of the emitting radicals, and especially of the effect of pressure on this temperature, seems to be one of the best ways of obtaining information about the method of formation of the excited radicals. Thus (Gaydon & Wolfhard 1948, 1949*a*) it was shown that the OH radicals in hydrocarbon flames have an effective rotational temperature of about 5400°K at atmospheric pressure, and this rises to still higher values at very low gas pressure. From this it could be concluded that in hydrocarbon flames the OH emission resulted mainly from chemiluminescence in which the radicals were formed in the electronically excited state and with high rotational energies, the very high effective rotational temperature being reduced from its initial value of at least 9000°K to 5400°K at atmospheric pressure by collisions with other molecules. For CH (Gaydon & Wolfhard 1949*b*) the effective rotational temperatures were always near the theoretical flame temperature and evidence for chemiluminescence was less definite. For C_2 one of us (Wolfhard 1939) has made some preliminary measurements at atmospheric pressure indicating a high effective rotational temperature, but detailed measurements with various gases at different pressures have not been made previously.

EXPERIMENTAL

The general principles of the method have been dealt with in the two previous papers in this series (1948, 1949*b*). For studying the Swan bands of C_2 it was necessary to use large dispersion, and a two-prism glass Littrow-type spectrograph with a dispersion of between 3 and $4\text{\AA}/\text{mm.}$ was employed. Ilford Press Ortho 2 plates were used, and calibration for change of plates sensitivity with wave-length was made by comparison with a tungsten strip-filament lamp running at known temperature, using Wien's law and data for the emissivity of tungsten at various wave-lengths and the required temperature. The photographic plates were calibrated with a rotating stepped sector and suitable source for spectrum lines (for some work an iron arc was used, while for many measurements it was found adequate to use the Hg line at $4916\text{\AA}$, the mercury arc being a steadier and more reproducible source).

The Swan bands are due to a $^3\Pi \rightarrow ^3\Pi$ transition, and the bands show only *P* and *R* branches. Each line is a close triplet due to the small spin splitting in the $^3\Pi$ states. Such a transition would normally show Λ -doubling, but in the case of C_2 , which is a homo-nuclear molecule, alternate lines are missing, so that in practice the appearance of the bands is normal with all lines present but without the Λ

doubling, since alternate components of the doubling are missing. Near the heads of the bands the structure was too complex for reliable intensity measurements owing to overlapping of lines, but away from the head the higher members of the R branch are fairly free from overlapping. For low rotational quantum numbers the triplet splitting due to spin was just resolvable, but at high rotational energy became too small to resolve. It was therefore decided to use a slit which was wide compared with this triplet splitting and to measure the three components together. About 25 lines of the (0, 0) band between values for K'' , the ground state rotational quantum number, of 42 and 97 were selected as being apparently free from overlapping and were used. The light from some of the flames studied was not very bright and in some cases it was only possible to record a dozen or so of these lines with the lower rotational energies.

A rotational analysis for the (0, 0) Swan band has been given by Shea (1927), but it is now known that his rotational quantum numbers are one unit too high, and the values adopted for these calculations were accordingly taken one lower. The relative transition probabilities for the lines of the R branch for a $^3\Pi-^3\Pi$ transition where both electronic states approximate closely to Hund's case b are proportional to $(2K''^2 + 4K'')/(K'' + 1)$. The actual intensities should then be proportional to

$$P\nu^4 \exp[-E_r/kT],$$

where P is the transition probability as given above, ν is the wave number of the line, k is the Boltzmann constant, T is the absolute temperature and E_r is the rotational energy of the upper level of the line. The values of E_r for the lines used were calculated from $E_r = B'K'(K' + 1) - D'K'^2(K' + 1)^2$, where the rotational energy constants have the values $B' = 1.7454$ and $D' = 6.8 \times 10^{-6}$. The effective rotational temperatures were then derived in the usual way by plotting

$$\log_e I - \log_e P\nu^4$$

against the rotational energy E_r , and obtaining T from the slope.

For the low-pressure flames the type of apparatus previously described was employed, burners of up to 54 mm. diameter being employed, giving pressures down to 2 mm. of mercury for oxygen/acetylene. For flames at atmospheric pressure an ordinary welding burner with a diameter of about 1.5 mm. was used. Gases were obtained from commercial cylinders without special purification.

In making the intensity measurements it was necessary to subtract the intensity of the continuous background from that of the lines. For this to be accurately measurable its intensity had to be well above the threshold sensitivity of the photographic plate, and this necessitated rather heavy exposures. Flames are, on the whole, rather feeble sources and hence long exposures were required. The chief limit to the accuracy of the results was found, in practice, to be temperature shifts in the spectrograph. It was found that a change of 1°C produced a serious shift and interfered with the resolution of the spectrum lines, and with hot flames running it was difficult to keep the laboratory temperature steady for more than an hour or so. Hence most of the measurements, and all the best values, were with the relatively bright acetylene flames. Useful results of rather lower accuracy were, however, obtained with some other gases.

RESULTS

The results for the various flames studied are summarized in table 1, in which the effective rotational temperatures as determined from the slope of the graph are given with the probable limits of error as estimated from the scatter of the points and the reliability of the plate. Two typical graphs are shown in figure 1 for stoichiometric acetylene/oxygen flames. There is, of course, an experimental scatter to the points, but in all plates these points lie around a straight line. This indicates that the distribution of rotational energy among the excited C_2 molecules follows closely a Maxwell-Boltzmann distribution law, and that we are therefore justified in assigning an effective rotational temperature to the molecules. Owing to the increasing spin splitting of the spectrum lines at low rotational quantum number it was not possible to make measurements below a rotational quantum number of 40, so we are unable to say whether this Maxwell-Boltzmann distribution holds at low rotational energy. In the case of OH it has been found (Gaydon & Wolfhard 1948) that the points lie on a fairly good straight line corresponding to a high effective rotational temperature at high rotational energy, but that there is usually some departure from a straight line at low rotational energies, the distribution of rotational energy among the lowest energy levels corresponding to a lower temperature which is near the true flame temperature. In the case of C_2 there is no sign of this over the rather considerable range covered by our observations (between a rotational energy of 3600 and 16,000 cm^{-1}), but such an effect might conceivably occur at very low rotational energy.

TABLE 1. EFFECTIVE ROTATIONAL TEMPERATURES OF FLAMES
FROM MEASUREMENT OF SWAN BANDS OF C_2

gas mixture	pressure	observed temp. (° K)
acetylene/oxygen: about stoichiometric	1 atm.	4950 ± 200*
slightly fuel-rich	1 atm.	4600 ± 500
rich	1 atm.	4700 ± 200
weak	1 atm.	4700 ± 200
stoichiometric, $\lambda = 1.00$	8.5 mm.	3800 ± 400*
rich, $\lambda = 0.70$	8 mm.	4200 ± 300
very rich, $\lambda = 0.50$	12 mm.	4200 ± 300
weak, $\lambda = 1.53$	9.5 mm.	3620 ± 300
stoichiometric, $\lambda = 1.00$	2.1 mm.	3800 ± 500
rich, $\lambda = 0.78$	2.0 mm.	4020 ± 400
acetylene/air: about stoichiometric	1 atm.	3400 ± 300*
rather rich	62 mm.	3240 ± 300
$1.33H_2 + 0.133 C_2H_2 + 1.00 O_2$ (stoichiometric)	9 mm.	3000 ± 500
acetylene/nitrous oxide: about stoichiometric	1 atm.	4020 ± 300
slightly rich	25 mm.	4150 ± 800
methane/oxygen: rich	1 atm.	3900 ± 800
rich, $\lambda = 0.52$	40 mm.	3250 ± 400
benzene/oxygen: stoichiometric, $\lambda = 1.00$	18 mm.	4150 ± 500
ethane/oxygen: nearly stoichiometric, $\lambda = 0.96$	25 mm.	4100 ± 400
stoichiometric, $\lambda = 1.0$	1 atm.	4100 ± 400
rich, $\lambda = 0.60$	1 atm.	3630 ± 400

* Mean value from the measurement of two or more plates.

All the effective rotational temperatures determined come appreciably above the calculated maximum flame temperatures for complete combustion. This is especially so for acetylene. The variation with mixture strength is not very marked; the highest temperatures are given by around the stoichiometric mixtures, but there is little change on the fuel-rich side even when there is considerable excess of fuel; weak mixtures do show some fall in effective rotational temperature.

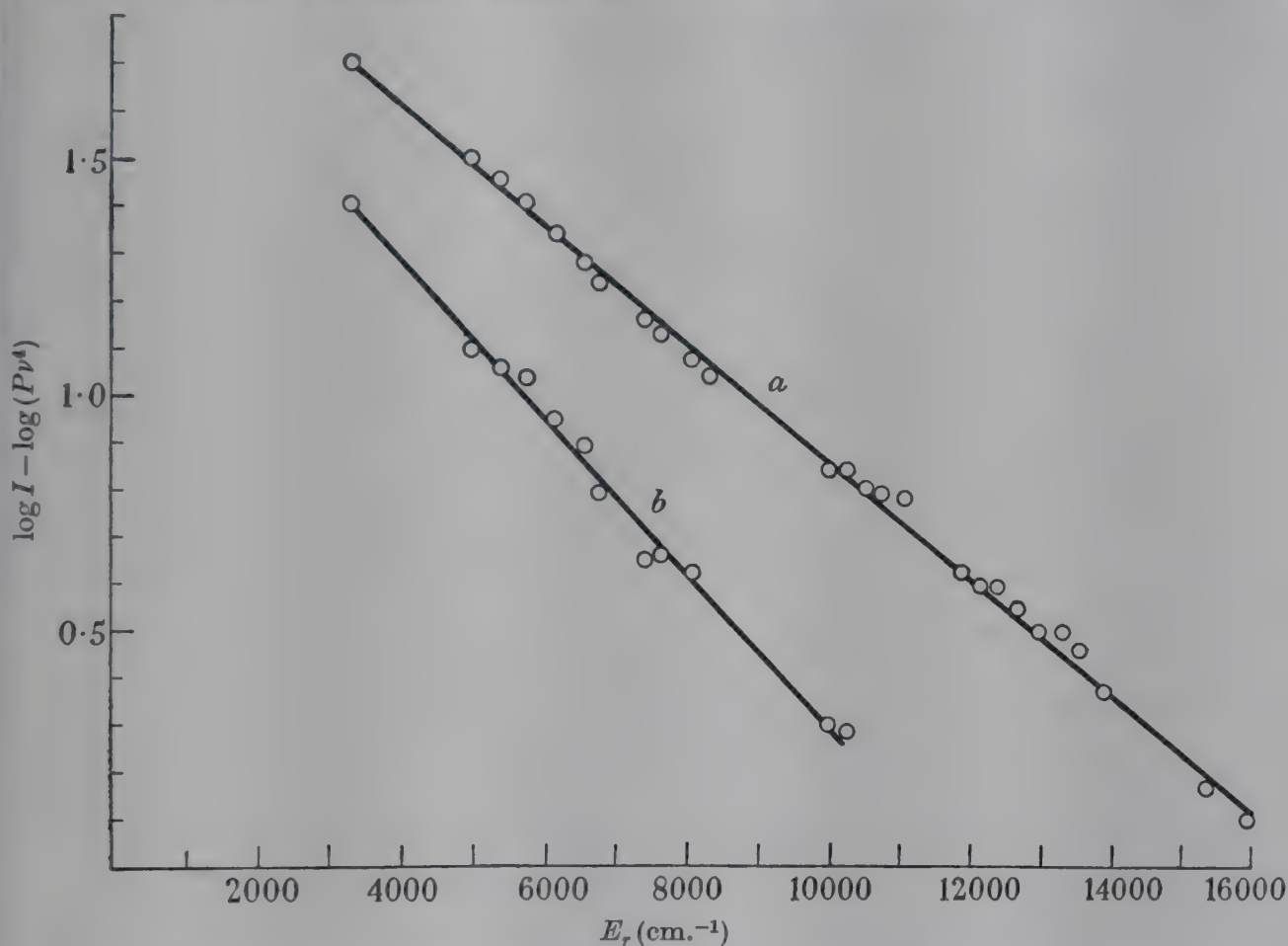


FIGURE 1. Graphs of $\log I - \log(P\nu^4)$ against rotational energy term, E_r , for stoichiometric oxygen/acetylene flames. (a) at 1 atm. with a welding burner, giving $T = 4880 \pm 200^\circ \text{K}$. (b) at 8.5 mm. with 21 mm. burner, giving $T = 3700 \pm 400^\circ \text{K}$.

There are three observations which seem of significance for the interpretation of the results. First, there is a definite fall in rotational temperature in the low-pressure flames; this is particularly marked for acetylene/oxygen where it falls from 4950 at 1 atm. to 3800° K at low pressure; this effect is quite real and outside experimental error as may be seen from the graphs in figure 1. This fall with pressure is also indicated for acetylene/air and the methane/oxygen, but not obviously for ethane/oxygen. The second observation is the lower temperature of acetylene/air compared with acetylene/oxygen; this again is well outside experimental error and shows that the effective rotational temperature of the C_2 bands is not solely determined by a simple chemical process between fuel and oxygen but is affected by diluents which lower the flame temperature. The third interesting observation is the relatively low value of about 3000° K for a flame of oxygen and hydrogen containing only a little acetylene; the true temperature of this flame will be a little lower than the pure

oxygen/acetylene flame but not so very much, and would be well above the temperature of the acetylene/air flame which gives a higher effective rotational temperature.

DISCUSSION OF THE METHOD OF EXCITATION OF C_2 IN FLAMES

The high values for the effective rotational temperatures at once suggest that we are dealing with a chemiluminescence. However, in this case it then becomes most difficult to explain the effect of pressure and the lower values for the flame with air or in the hydrogen flame.

For OH we found a rather high effective rotational temperature around 5400°K for all hydrocarbon flames at atmospheric pressure, rising to a much higher value at low pressure (Gaydon & Wolfhard 1948). This effect of pressure was satisfactorily explained by assuming chemiluminescence accompanied by deactivation processes which were less important at low pressure owing to the small number of collisions during the radiative life of the excited molecules. For C_2 it is not possible to give a similar explanation because the temperature falls with pressure instead of rising.

Also, if we were dealing with chemiluminescence we should expect the effective rotational temperature to be fixed by the chemical reaction and inert diluents such as nitrogen (in the air flame) should not affect the rotational temperature. Also we have not succeeded in suggesting a plausible mechanism giving sufficient energy for the chemiluminescence.

In the earlier paper (1948) we also showed that excitation temperatures, as measured by the spectrum-line reversal method, were abnormally high in the reaction zones of flames. Using iron lines near 2500°A we found an excitation temperature around 3300°K for an acetylene/air flame at low pressure, while for acetylene/oxygen the temperature was still higher. It thus seems likely that the effective rotational temperatures for C_2 are related, perhaps indirectly, to the excitation temperatures.† We know that the excitation temperature for acetylene/oxygen is much higher than that for acetylene/air, and that the effective rotational temperatures are in the same order. It is likely that the excitation temperature for acetylene flames is higher than for other hydrocarbon flames, and not unreasonable to expect that it will fall with gas pressure. Oxygen/hydrogen flames do not show the well-marked reaction zone (inner cone-type) like hydrocarbon flames, and the excitation temperature probably coincides fairly well with the true flame temperature; in oxygen/hydrogen flames containing a little acetylene we may expect excitation temperatures only a little above the flame temperature; we have seen that the rotational temperature is only a little above the expected flame temperature. Thus it seems reasonable to assume that there is some relation between excitation temperature and effective rotational temperature of C_2 .

We are not aware of any detailed explanation of the high excitation temperatures in the reaction zones of hydrocarbon flames, and it is desirable to discuss the subject here. The energetic particles responsible for the abnormally high excitation can hardly be fast-moving atoms or molecules because, first, this excess energy would

† The brightness temperature for C_2 bands in an acetylene/oxygen flame at 1 atm. is known to exceed 4000°K ; see following part.

very quickly be lost by collision with other molecules, and secondly, because it seems that the chance of electronic excitation, i.e. transfer of a relatively light electron from one orbit to another, by collision with a relatively massive particle possessing translational energy only is extremely small. In pure flame gases the number of free electrons is very small and could hardly account for the strength of emission of light from flames, even if a high proportion of the electrons possessed abnormally high kinetic energy. This leaves two reasonable possibilities, either the active particles are endowed with excess electronic energy (i.e. they are electronically excited as the result of chemical action), or they possess excess vibrational energy. In either case the transfer of this energy to other molecules might occur with reasonable efficiency, but it would not be dissipated quickly by conversion into translational energy. It is likely that in the combustion of CO electronically excited molecules of CO_2 with a relatively long radiative life are formed, and also we have seen previously (Gaydon & Wolfhard 1948) that in the reaction zone of hydrocarbon flames excited OH radicals are formed by chemiluminescence. In the case of excited CO_2 it seems difficult to explain the strength of the C_2 bands at the very base of the reaction zone if they were excited by activated CO_2 . In the case of OH, the 3064 Å band is also strong at the base of the reaction zone and throughout this zone, and those flames which give strong C_2 bands all show strong OH chemiluminescence, but it is doubtful whether the electronically excited OH could account quantitatively for the C_2 emission; the excited OH are, on the average, deactivated after around 100 collisions,[†] and as the C_2 concentration is low[‡] the number of photons emitted by C_2 should be several orders lower than from OH, whereas in rich flames it seems that the C_2 radiation is as strong as or stronger than the OH radiation. Also we have observed abnormally strong excitation of Pb and Fe lines requiring a higher energy for their excitation than could be supplied by electronically excited OH. While the possibility of other electronically excited species being present cannot be entirely eliminated, we are of the opinion that the most likely cause of the high excitation in flames is the presence of molecules with excess vibrational energy.

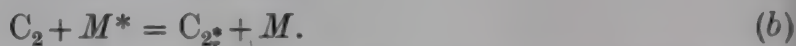
It is known from experiments on supersonic dispersion that vibrational energy may be retained through a large number of collisions and is not readily transformed into translational energy (for references see Gaydon 1948). On the other hand, experiments on quenching of fluorescence show that electronically excited atoms or molecules cannot be deactivated at all readily by monatomic gases, but may be deactivated fairly well by diatomic gases and very readily by polyatomic molecules possessing a large number of vibrational degrees of freedom. Presumably the reverse process may occur, and diatomic or polyatomic molecules with high vibrational energy could cause electronic excitation on collision. It was suggested long ago

[†] In the previous paper it was shown that if we assumed a normal collision cross-section for excited OH, then the electronically excited radicals were deactivated on the average after about 40 collisions; later work suggests that the collision cross-section for the excited radical may be rather higher so that it is safer to assume a figure of about 100.

[‡] The low concentration of C_2 is shown by failure to detect self-absorption of C_2 bands (Wolfhard 1939); these experiments to detect self-absorption have been repeated, with negative results, for the low-pressure flames using rich mixtures and a flat flame giving an absorption path of several cm.

(Egerton 1928) that in the combustion of hydrocarbons, energetic hydrocarbon molecules combine momentarily with oxygen to form a temporary peroxide in a high energy state, the subsequent breaking down of the peroxide setting free the energy to maintain the combustion process. While we have no direct evidence that it is these energy-rich peroxides which are responsible for the high excitation temperatures in flames it remains a reasonable possibility. Certainly there may be a tendency in flames for the formation of short-lived collision complexes which retain their energy of formation as excess vibrational energy.

We still require to know something about how the C_2 molecules acquire the high rotational energies which are responsible for the observed high rotational temperatures. There are two possibilities to consider; either the C_2 molecules possess the high rotational energy from the start, or they acquire it at the moment of excitation by the vibrational-energy-rich molecule. Writing C_{2r} for a C_2 radical with excess rotational energy, and C_{2*} for electronically excited C_2 with excess rotational energy as well, the possibilities may be written



Case (b) requires the conversion of a fair proportion of the energy of the energy-rich body M^* into rotational energy; if M^* is a relatively large polyatomic molecule it seems rather unlikely that the C_2 radical could gain a lot of rotational energy during the collision because of the difficulty of conserving the angular momentum of the system. We should expect that on diluting a flame with nitrogen or hydrogen the proportion of energy-rich molecules would change, but that the character and energy of such molecules would remain unaltered, and hence, it is not easy to see for case (b) why the acetylene/air flame or the acetylene/hydrogen/oxygen flame shows so much lower rotational temperatures than the acetylene/oxygen flame. If, however, the C_2 radicals are normally formed in some exothermic reaction or decomposition and possess, to begin with, high rotational energy (case a) then this energy will gradually be lost by collision with inert molecules and the effective rotational temperature at the moment of electronic excitation will depend on how many collisions with inert molecules occur before the C_2 meets the energy-rich exciting molecule. Thus the extent to which the rotational energy falls will depend on the concentration of energy-rich molecules, that is, it will be related to the excitation temperature. Hence, dilution with nitrogen or hydrogen could bring down the effective rotational temperature.

We may summarize this rather long discussion by saying that the high effective rotational temperature of the C_2 radicals in flames, and the effect of pressure and dilution on it, is best accounted for by assuming not true chemiluminescence, but an abnormally high excitation in the flame due to molecules, perhaps of peroxidic character, possessing excess vibrational energy. The C_2 radicals may be formed with high rotational energy in some exothermic reaction or decomposition, and their average rotational energy depends on the number of collisions with inert molecules suffered before excitation, and hence depends on the effective excitation temperature of the flame.

The idea that molecules with excess vibrational energy are present in flames may also have a bearing on the problem of ionization in flames. Calcote (1949) has summarized evidence that flame gases may be more highly ionized than calculations for equilibrium mixtures would suggest, and that in the inner cone (i.e. the reaction zone) there may be as many as 10^{12} ions/cm.³. If the energy-rich molecules are collision complexes, then we may expect them to have a short life and to decompose spontaneously if allowed sufficient time. The decrease in the number of active molecules by spontaneous decomposition may account for the observed fall in effective rotational temperatures at low gas pressure. It should be accompanied by a fall in the excitation temperature of the flame. Unfortunately, measurements of excitation temperatures are complicated by the fact that they often exceed the temperature of the hottest available background source (the positive crater of the carbon arc at 3800° K), and hence straightforward measurements by the spectrum-line reversal method cannot be made. We have hopes of devising a method of measuring excitation temperatures which will get around this difficulty.

In conclusion, it is a pleasure to thank Sir Alfred Egerton, F.R.S., for his interest in this work, and one of us (A.G.G.) is grateful to the Royal Society for the award of a Warren Research Fellowship.

REFERENCES

- Calcote, H. F. 1949 *Third symposium on combustion*, p. 245, Wisconsin, Baltimore: Williams and Wilkins.
- Egerton, A. C. 1928 *Nature*, **121**, 10.
- Gaydon, A. G. 1948 *Spectroscopy and combustion theory*, 2nd ed. London: Chapman and Hall.
- Gaydon, A. G. & Wolfhard, H. G. 1948 *Proc. Roy. Soc. A*, **194**, 169.
- Gaydon, A. G. & Wolfhard, H. G. 1949a *Third symposium on combustion*, p. 504. Wisconsin, Baltimore: Williams and Wilkins.
- Gaydon, A. G. & Wolfhard, H. G. 1949b *Proc. Roy. Soc. A*, **199**, 89.
- Shea, J. D. 1927 *Phys. Rev.* **30**, 825.
- Smith, E. C. W. 1940 *Proc. Roy. Soc. A*, **174**, 110.
- Wolfhard, H. G. 1939 *Z. Phys.* **112**, 107.

Spectroscopic studies of low-pressure flames

IV. Measurements of light yield for C_2 bands

BY A. G. GAYDON AND H. G. WOLFHARD

Chemical Engineering Department, Imperial College, London

(Communicated by Sir Alfred Egerton, F.R.S.—Received 29 October 1949

—Read 2 February 1950)

The light yield for the Swan bands is measured for a number of hydrocarbons. For $C_2H_2 + 2\frac{1}{2}O_2$ flames 3×10^{-5} photons are emitted for each O_2 consumed at 1 atm. pressure, rising to 13.5×10^{-5} at 3.5 mm. pressure. For propane the effect of pressure is more marked. The effect of mixture strength and of diluents is studied. For partial replacement of hydrocarbon by H_2 or CO (the mixture being held stoichiometric) the yield falls with surprising rapidity, being proportional to a high power of the hydrocarbon concentration (square to sixth power). Preliminary results for CH are similar. The C_2 emission does not appear to be due to true chemiluminescence, but there is abnormally high excitation by some active species in the flame, and the concentration of C_2 exceeds the values found by Brewer, Gilles & Jenkins for equilibrium with solid carbon at the flame temperature. Polymerization and formation of unsaturated compounds appears to be important. The relation between C_2 emission and the formation of solid carbon is discussed; polymerization of C_2 cannot account quantitatively for the formation of solid particles.

INTRODUCTION

In previous papers (Gaydon & Wolfhard 1948, 1949*b*, 1950) we have stressed the advantage of using flames at very low gas pressure for spectroscopic study, have described the experimental technique for running these flames, and have given various results, including measurements of effective rotational temperatures for OH, CH and C_2 bands. It has been shown that the main part of the OH radiation from hydrocarbon flames is due to a true chemiluminescence, but that for the 4315 and 3900 Å bands of CH and the Swan bands of C_2 it is more likely that the radicals are formed in the normal electronic state and subsequently excited by collision with energy-rich molecules formed in the combustion processes. In an attempt to track down the method of formation and excitation of these radicals in flames we have made measurements of the light yield from flames (expressed in photons emitted per molecule consumed in the combustion), and especially of the effect of various conditions such as pressure, nature of the fuel, mixture strength, and effect of diluents or partial replacement with another fuel on this light yield. This paper is concerned with the light yield of the Swan bands of C_2 , for which measurements have been made for a number of fuels with a variety of conditions. Some preliminary results for CH are also referred to.

A few measurements have been made at atmospheric pressure and at intermediate pressures, but most of the results are for the very low pressure flames, around $\frac{1}{100}$ atm. There are several advantages in the study at low pressure. The formation of solid carbon, which renders the flame luminous and unsuitable for spectroscopic study, is reduced or even eliminated at low pressure, so that it is possible to measure the C_2 emission even from flames of very rich mixtures. The reaction zone, or flame front, is greatly enlarged in all dimensions at low pressure, and the emission from this region compared with that from the hot interconal gases is much greater, while the

emission from the outer cone is entirely eliminated in the closed vessels used for the low-pressure flames. The general character of the combustion processes and the radiation at low pressure appears to be essentially similar to that at atmospheric pressure, but there may be a tendency for the reactions to be simplified at low pressure because two-body collisions will be favoured at the expense of three-body or multi-body collisions. Also any true chemiluminescence may tend to show up more at low pressure because the number of collisions during the radiative life of the excited radical will be much fewer and there will be less chance of quenching. The continuous background, present in all flame spectra and attributed to recombination of ions with free electrons, is much weaker at low pressure.

EXPERIMENTAL ARRANGEMENT

As it was required to measure the intensity of the whole reaction zone of the flame, to express the results in number of photons emitted per molecule consumed, rather than the brightness of the flame per unit area, it was necessary to use a spectrograph alone, without any lens to condense the light on to the slit. The spectrograph had to be of adequate aperture and at adequate distance to collect light from the whole of the flame. The spectrograph was a quartz prismatic instrument made by us; it gave about the same dispersion as commercial small quartz instruments, but had a relatively large aperture ($f/4.5$); for all these measurements it was set with the slit at exactly 1 m. from the centre of the flame. A relatively wide slit of about 0.1 mm. was used; this enabled the heads of the vibrational structure to be distinguished easily, but did not of course separate the rotational fine structure.*

The intensity of the light from the flame, at various wave-lengths, was compared with that from a strip-filament tungsten lamp running at known brightness temperature. The lamp was set at a measured distance between the flame and the slit (usually at about 80 cm. from the slit), and a length of the centre of the strip filament was selected with a suitable stop. From a knowledge of the width and distance of this stop and the width of the filament the effective area of filament exposed to the spectrograph could be calculated. A rotating sector of variable angle was placed just in front of the spectrograph slit. By taking a spectrogram of the flame with suitable sector setting and a number of spectra (on the same plate and with identical exposure times) of the tungsten lamp with different sector settings it was possible to compare the intensity of the flame and lamp. A correction, using the inverse square law, was of course necessary to allow for the different distance of the flame and lamp from the slit. The emission of the lamp was calculated from its brightness temperature (determined from measurement of the current with the aid of a National Physical Laboratory calibration) and the known emissivity of tungsten at various wave-lengths. In practice it was possible to take fourteen exposures on a plate, usually six sector settings on the lamp and up to eight exposures on different flames. The brightness temperature of the lamp was adjusted to give an intensity comparable with that of the flame. For most of the work the current through the lamp was

* The measurements on light yield were shown to be unaffected by self-absorption in the flame. Attempts to detect absorption by reflecting an image of the flame back through the flame were always negative.

12.00 Å, giving a brightness temperature of 1947° K at 6500 Å. Exposure times for various plates ranged from 20 sec. to 5 min.

For the low-pressure flames the type of burner previously described was used. The largest burner diameter suitable for this work was 35 mm., as with larger burners the light from the flame became partially obscured by the edges of the quartz window which was 6 cm. across. Thus the lowest pressure attained was about 3 mm. with oxygen/acetylene. For most of the measurements, burners of 25 or 28 mm. diameter were used. For the work at atmospheric pressure an ordinary welding burner with a suitable nipple of about 1 mm. diameter was employed.

Gases were measured with flow-meters in the usual way, the flow-meters being calibrated with a gas meter. Changes in room temperature caused slight errors due to change in viscosity. These were probably the most serious errors in the experiments as errors of up to 2 % could occur, and these might influence the mixture strength, to which the light yield was very sensitive. For vapours and gases such as acetylene which could not be measured with the gas meter the calibration was made by comparing the rate of pressure rise in the vessel when the pump was turned off with that for a known flow of oxygen or other gas which had been metered.

The gases used were of ordinary commercial purity except the butane, which was specially supplied by Imperial Chemical Industries, and the methane which was purified to at least 99.8 % by Petrocarbon Ltd.

EFFECT OF MASS FLOW, FLAME SHAPE AND PRESSURE ON LIGHT YIELD

In order to compare the light yield from different fuels and with various mixture strengths, diluents, etc., it was necessary as a preliminary to examine the influence of the shape and character of the flame and the effect of pressure on the radiation, as different gas mixtures would give flames of different character and at different pressures if the same burner diameter was used. These preliminary measurements were mostly made with stoichiometric oxygen/acetylene mixtures, as these gave relatively bright flames and could be maintained over the greatest pressure range. The relative intensity of the (0, 0) band of C_2 was measured, but results are expressed as absolute light yield per O_2 molecule consumed by comparison with the measurements on the absolute light yield described in the next section.

With the apparatus used for the low-pressure flames the mass flow of fuel and oxygen could be set at any required value, and then the pressure in the combustion vessel could be varied by adjusting the rate of flow of gas to the pump. For a fixed burner diameter there was always a certain minimum mass flow below which a flame could not be maintained at any pressure. At just above the minimum mass flow a comparatively flat disk-shaped flame was obtained. With a little higher mass flow the shape could be varied slightly, being flat at the higher pressure limit, but rising and becoming conical at lower pressure. At still greater mass flow the shape varied from nearly flat to a good cone of the Bunsen burner type. With high mass flow, however, the flame became unsteady owing to turbulence, and it was necessary to work at relatively high pressure to prevent blow-off; also the vessel tended to become too hot with very big mass flows.

It was found that for any fixed burner diameter and a suitable mass flow, fairly well above the minimum value, changing the shape of the flame from flat to a cone about twice as high as the burner diameter by slight variation in gas pressure did not alter the amount of C_2 radiation beyond the limits of experimental error.

With fixed burner diameter and varying mass flow it was found that the change in light yield with mass flow was slight. The light yield fell off somewhat very near the limiting mass flow, probably because of loss of heat and active radicals to the surroundings. Except near the minimum mass flow the increase in light yield with mass flow was small. It was therefore assumed that provided reasonably high mass flows were used the light yield was unaffected by small changes.

TABLE 1. EFFECT OF PRESSURE AND MASS FLOW ON LIGHT YIELD.
STOICHIOMETRIC OXYGEN/ACETYLENE

burner diameter (mm.)	O ₂ flow, (ml./sec. at N.T.P.)	pressure (mm. Hg)	light yield in quanta/O ₂ molecule ($\times 10^{-5}$)
(a) at constant mass flow/burner diameter			
35.1	33.4	3.5	13.5
33.6	29.9	3.4	13
28	26.0	4.5	12
28	26.0	5.0	12.6
21	20.0	5.5	11.3
13	12.4	8.0	7.4
7.7	7.35	17	6.9
3.5	3.3	28	4.6
(b) at relatively higher mass flow			
7.7	46.4	25	7.9
3.5	7.35	36	7.4
2.25	3.3	74	4.9
2.25	7.35	84	4.0
0.97*	20	760	3.6

* Welding burner.

Thus for all the measurements on the low-pressure flames described here, a mass flow fairly well above the minimum value was used, and as far as possible a conical flame about as high as the burner diameter was maintained. Under these conditions any errors due to change of light yield with mass flow appeared negligible.

It has been established that the minimum pressure at which a flame can be maintained is approximately inversely proportional to the diameter of the burner, and that the minimum mass flow is very nearly proportional to the burner diameter. Thus for comparing flames at different pressures a series of burners with diameters ranging from 35 to 3.5 mm. were used, the mass flow being adjusted to be proportional to the burner diameter. For smaller burners (at higher pressures) it was impracticable to decrease the mass flow proportionately to the burner diameter because the flame became too weak and small to be photographed in reasonable time, and hence relatively higher mass flows were used; these may give a rather higher light yield on this account. Results are summarized in table 1.

From these results it is clear that the light yield decreases as the pressure is raised. Unfortunately, results are not strictly comparable over the whole range. Over the eight-fold pressure rise from 3.5 to 28 mm. the light yield falls nearly to a third, but at higher pressure the falling off may be less rapid. Since the light yield at the greater mass flows is somewhat higher, results are not strictly comparable. For the welding flame the relative mass flow used was very high, and the measurement was complicated, as previously pointed out, by radiation from the outer cone of the flame; there appeared to be an appreciable amount of continuous spectrum and radiation characteristic of burning CO. By correction for this background the light yield might be brought down to about 3×10^{-5} C₂ quanta per O₂ molecule consumed, but it is known (Gaydon & Wolfhard 1950) that the effective rotational temperature of the C₂ bands is higher at atmospheric pressure, and it is uncertain how much of the extra background near the band is due to CO-flame spectrum and how much results from the extended rotational structure of the bands.

For oxygen/acetylene the effect of pressure on light yield for rich mixtures cannot be studied fully because of the formation of solid carbon at high pressure. Over a range of pressures which is not too high the light yield again falls with increasing pressure.

For propane some measurements have been made at low and at atmospheric pressure for both stoichiometric and rich mixtures. Results are summarized in table 2. For these measurements the oxygen flow was always 26 ml./sec. at N.T.P. We have previously seen that there is some tendency for the light yield to increase at higher ratios of mass-flow to burner-diameter, so the results are again not strictly comparable, but serve to show the order of the decrease in light yield at higher pressure for the stoichiometric and rich mixture.

TABLE 2. EFFECT OF PRESSURE ON LIGHT YIELD FOR PROPANE/OXYGEN

mixture	burner diameter	pressure	light yield in quanta/O ₂ molecule used
	(mm.)		($\times 10^{-5}$)
C ₃ H ₈ + 5O ₂	21.	12 mm. Hg	0.78
	1.4	1 atm.	0.040
rich*	21	24 mm.	6.0
	1.4	1 atm.	0.40

* The rich mixture was $\lambda = 0.55$, i.e. the mixture was (1/0.55) parts of propane to 5 of O₂. For the rich mixtures values were interpolated from the light yield versus mixture strength curves given later, the maximum C₂ light yield occurring at $\lambda = 0.55$ for propane.

It can be seen that, making some allowance for the different pressures of the stoichiometric and rich mixtures, the decrease in light yield with higher pressure is about the same for the two mixtures. For the stoichiometric mixture the light yield falls nearly to 1/20 over the 63 to 1 rise in pressure, without correcting for the effect of mass flow; if this were taken into account the change would be even greater. The effect of pressure on light yield is therefore greater for propane than for acetylene.

MEASUREMENT OF THE ABSOLUTE LIGHT YIELD

The spectrum of a stoichiometric oxygen/acetylene flame was compared quantitatively with that of a known area of the tungsten strip-filament lamp. The flame was maintained at a burner of 21 mm. diameter and pressure of 5.5 mm. Hg. The oxygen flow was 20 ml./sec. and the acetylene flow 8 ml./sec. The energy radiated by the lamp, J_λ , in watts per steradian is given by

$$J_\lambda = \frac{Ac_1\lambda^{-5}}{e^{c_2/\lambda T} - 1} \cdot E_\lambda d\lambda$$

where λ is wave-length in cm., A is area in cm.², T is absolute temperature, E_λ is emissivity of tungsten, $c_1 = 1.177 \times 10^{-12}$ W cm.², $c_2 = 1.438$ cm.deg. and $e = 2.718$. The flame radiates over a whole sphere, and the total energy radiated by the flame, must be compared with $4\pi J_\lambda$.

Using the above formula and the measured area and brightness temperature of the lamp (brightness temperature 1947° K, giving a true temperature of 2098° K) curves of optical density of the photographic plate were plotted against log intensity at intervals of 100 Å from 4500 to 5800 Å, this covering the region of the (1, 0), (0, 0) and (0, 1) sequences of C₂. Using these curves and measurements of the plate density for the flame spectrum at intervals of 25 Å a curve of intensity against wave-length was then plotted for the flame. A graphical integration of the area under this curve then enabled the total radiation from the flame, in watts, to be calculated. The value obtained was 18×10^{-3} W; this is for radiation in the (1, 0), (0, 0) and (0, 1) sequences only. The radiation in these three sequences was nearly the same. For the other sequences the radiation was very much less; the (0, 2) band came rather far in the red for accurate measurement and the (2, 0) sequence was masked by the CH band. A value of 21×10^{-3} W was adopted for the total radiation from the flame. A correction of 10 %, bringing the value to 23×10^{-3} , was applied for loss of light by reflexion in the two surfaces of the quartz window.

To express the result in photons per molecule consumed in the flame it is necessary to use the usual Planck relation $E = h\nu$, and to find the number of molecules entering the flame per sec. from the gas flow and Avogadro's number. In order to be able to compare the emission from different fuels it seemed best to express the result in number of photons per oxygen molecule consumed. The number of photons per second emitted by this flame was 6.0×10^{16} , and the number of oxygen molecules entering the flame was 5.4×10^{20} per sec., giving a light yield of 11.3×10^{-5} quanta per O₂ molecule.

It was obviously not practicable to carry out a detailed calculation of this nature for a number of fuels under a great variety of conditions, and other measurements were made relatively by comparing the intensity in the (0, 0) band of C₂. There is no obvious change in the relative intensity of the three main band sequences of C₂ from one flame to another; we know from previous work that the effective rotational temperature does change somewhat, and this may affect the distribution of intensity within the band, but errors due to variation in the vibrational and rotational intensity distribution are likely to be small and probably not greater than those due to

the limitations of the photometry and measurement of gas flow; with the small resolving power and relatively wide slit the region of the (0, 0) band photometered covered most of the lines of the *R* branch which forms the band head.

EFFECT OF MIXTURE STRENGTH AND NATURE OF FUEL

The influence of the mixture strength has been determined for flames of several hydrocarbons burning with oxygen at low pressure. The results are expressed in photons emitted by C_2 per O_2 molecule consumed. The mixture strength is expressed by the coefficient $\lambda = (\text{actual } O_2 \text{ flow})/(\text{theoretical } O_2 \text{ flow for a stoichiometric mixture})$. The procedure was to take a certain stoichiometric mixture as basis and on the rich side to decrease λ by increasing the fuel flow, and on the weak side to increase λ by increasing the O_2 flow; it was assumed in calculating the light yield that the amount of O_2 consumed was always that required for the stoichiometric mixture, the excess in weak mixtures merely acting as diluent.

For acetylene, measurements were made using burners of 34, 21 and 7 mm. diameter, and results are shown in figure 1. Apart from the influence of pressure, the three curves are similar, showing a sharp maximum near $\lambda = 0.42$, i.e. a very rich mixture containing more than twice the theoretical amount of C_2H_2 . Slight carbon formation commenced at about $\lambda = 0.33$ for the 21 mm. burner and at about 0.35 for the 7 mm. burner. The curves shown in figure 1 are uncorrected for the effect of pressure, the pressure changing, of course, with mixture strength. Any correction would strengthen the already sharp maximum.

Rather similar results were obtained for the other hydrocarbons, although the maximum at rich mixtures was usually less sharp. Table 3 gives the results.

TABLE 3. LIGHT YIELD IN QUANTA PER O_2 MOLECULE FOR STOICHIOMETRIC (*st.*) AND RICH MIXTURES

fuel	light yield (<i>st.</i>) ($\times 10^{-5}$)	pressure <i>st.</i> mixture (mm. Hg)	light yield at maximum ($\times 10^{-5}$)	λ for maximum	pressure at maximum (mm. Hg)
CH_4	0.049	13	1.05	0.59	30
C_2H_2	11.3	5.5	112	0.42	14
C_2H_4	2.9	8.5	16	0.58	18
C_2H_6	1.4	15	4.6	0.64	26
C_3H_8	0.78	12	6.1	0.55	24
C_6H_6	14.3	7.6	33.5	0.66	16

Acetylene and benzene give a much higher light yield than any other fuels. The results for ethane and propane show rather unexpected differences. For other fuels such as alcohols, ethers and aldehydes the C_2 light yield is very much less, especially for the lower members of the series, but no measurements are available, as the flames were too weak for measurement with the present arrangement. Preliminary measurements of the effect of mixture strength on the relative light yield for the CH bands shows a similar trend to that for C_2 , but the maximum lies at less rich mixtures ($\lambda = 0.59$ for C_2H_2 , $\lambda = 0.85$ for C_6H_6) and the strength of the radiation does not fall off nearly so rapidly for weak mixtures.

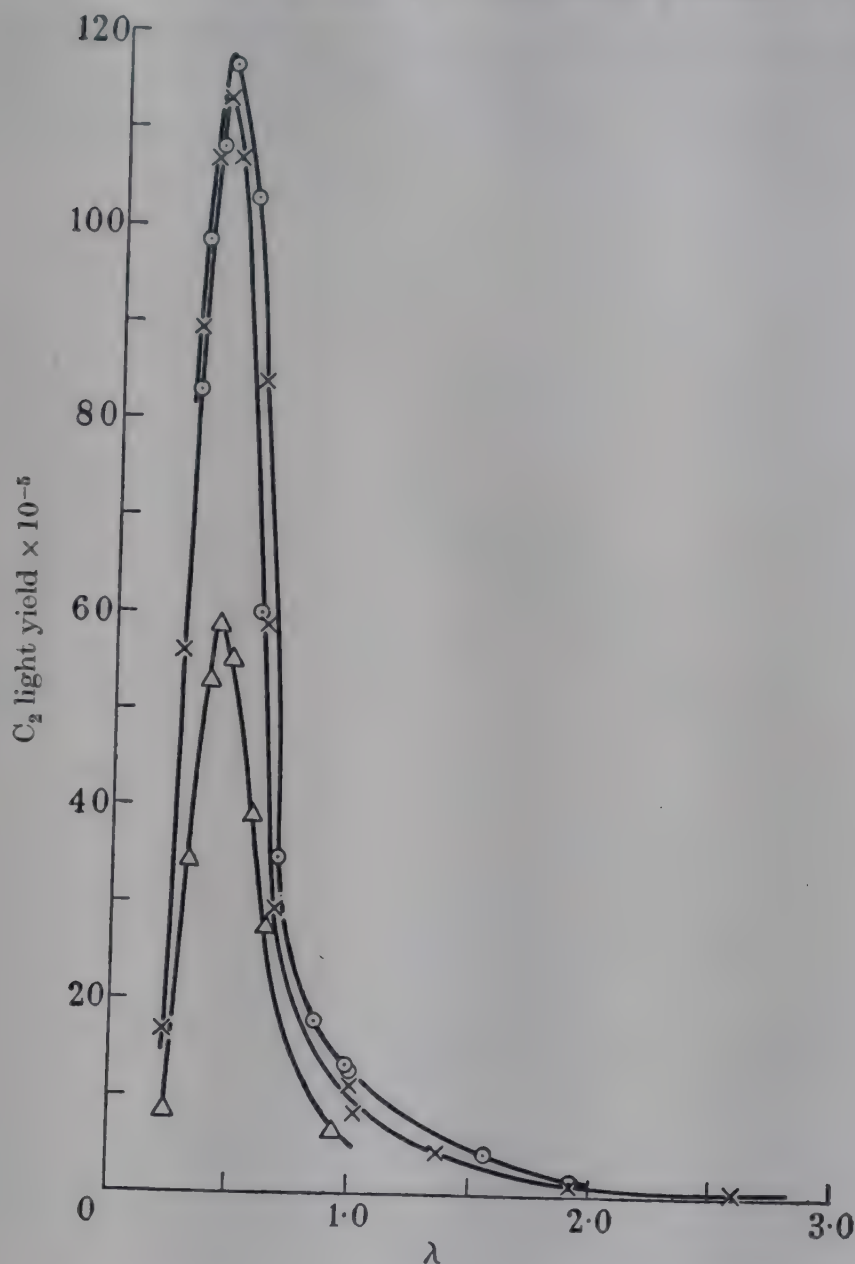


FIGURE 1. Effect of mixture strength on light yield for oxygen/acetylene flames.
 Sizes of burner: $\odot$, 34 mm.; $\times$, 21 mm.; $\triangle$, 7 mm.

EFFECT OF DILUENTS

The C_2 light yield has been compared for a stoichiometric oxygen/acetylene flame burning at low pressure on a 28 mm. burner with the addition of varying amounts of various diluents. Results are shown in figure 2. The amount of diluent is expressed by the constant n for a mixture $C_2H_2 + 2\frac{1}{2}O_2 + nX$, where X is the diluent.

Nitrogen causes a moderate weakening in the light yield. The fall in excitation temperature on adding the diluent and the rise in the pressure at which the flame is maintained would appear to be adequate to account for the observations, and there is no evidence that N_2 molecules have any marked quenching action. The interesting features are the very marked decrease in light yield on adding oxygen, and even sharper increase on adding more acetylene; these changes correspond, of course, to altering the mixture strength. It seems that acetylene promotes the formation or excitation of the C_2 while oxygen inhibits it. The moderate increase in light yield

on adding hydrogen and the very slight initial increase with carbon monoxide may be attributed to the effect of these gases in reducing the equilibrium concentration of oxygen rather than to any specific action on their own. A very similar series of results was obtained for ethane.

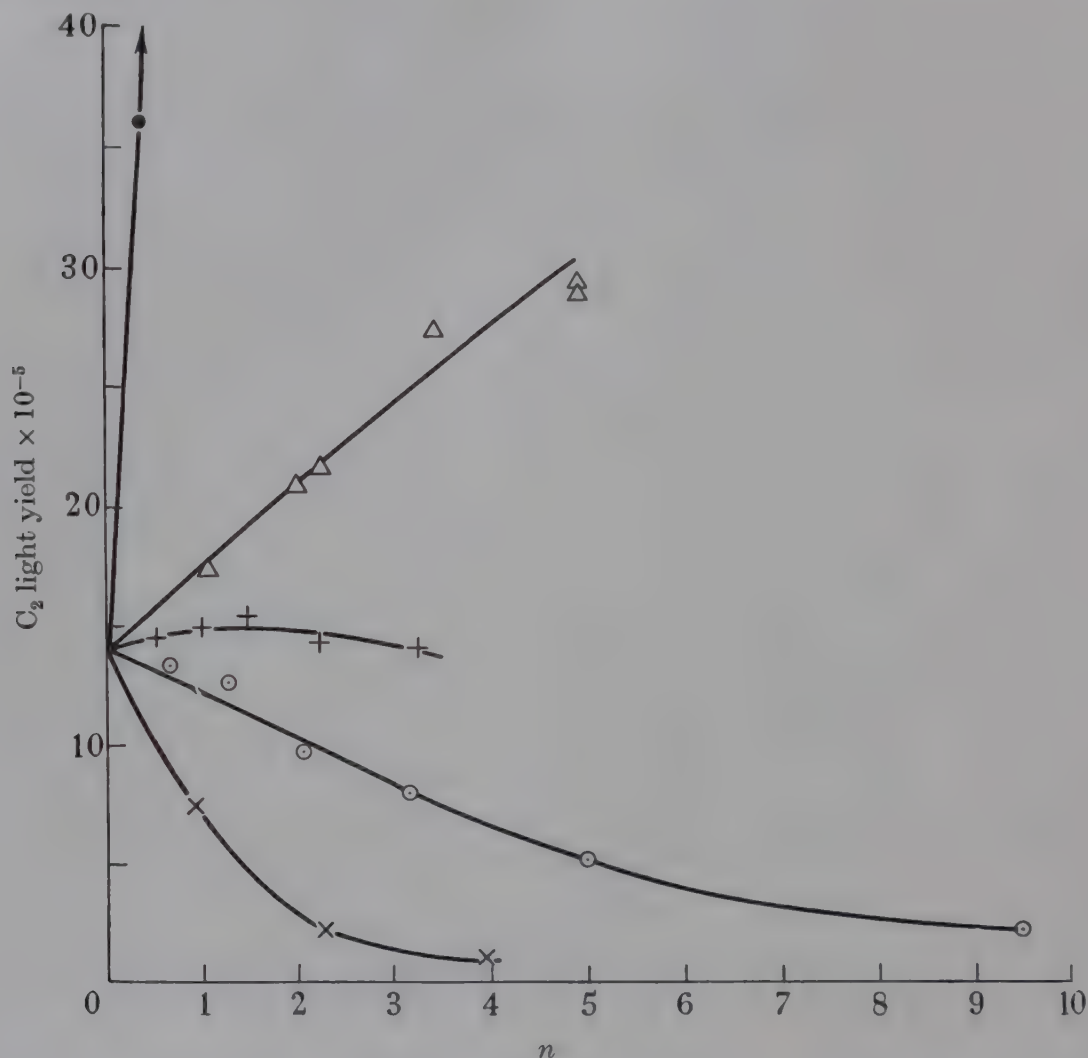


FIGURE 2. Effect of diluents on light yield from oxygen/acetylene flame.

●, C₂H₂; △, H₂; +, CO; ○, N₂; ×, O₂.

EFFECT OF REPLACEMENT WITH H₂ AND CO

Some most interesting results were obtained by measuring the light yield for stoichiometric mixtures in which the hydrocarbon was progressively replaced by either hydrogen or carbon monoxide in amounts to keep the mixture stoichiometric. Results for ethane are shown in figure 3. For the replacement of ethane by hydrogen the pressure at which the flame is maintained can be kept almost constant (at between 11 and 12 mm. Hg) and no big change in theoretical flame temperature is to be expected, so that effects of pressure and temperature should be negligible. One might expect, taking a simple view, that the light yield would decrease proportionately to the ethane concentration. Actually it falls off much more rapidly as the ethane is replaced with hydrogen, the light yield varying rather more rapidly than as the square of the ethane concentration. For replacement with carbon monoxide the observed light yield falls off even more rapidly, being proportional to

rather more than the fourth power of the ethane concentration. In this case there is some rise in the pressure as we pass to a higher proportion of CO, so the results are not so strictly comparable. Also, the spectrum shows some of the continuous and

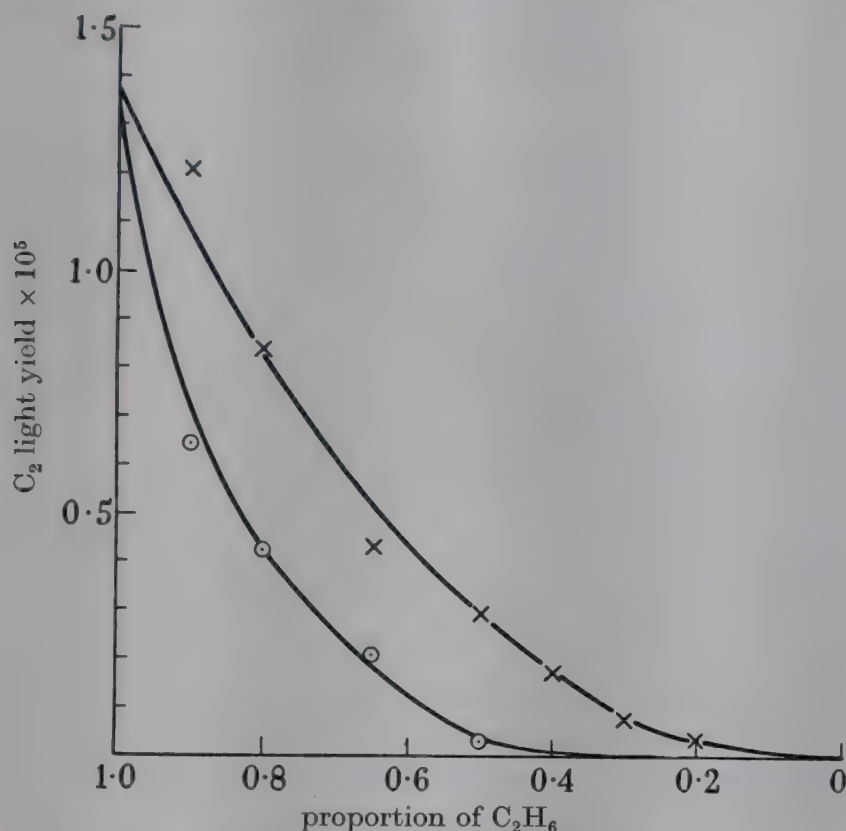


FIGURE 3. Light yield for stoichiometric ethane/oxygen flames in which ethane is replaced by $\times$, H_2 or $\circ$, CO.

slightly banded structure characteristic of burning CO for mixtures containing the higher proportions of CO, so that it became necessary to estimate the strength of this background spectrum and subtract it from the intensity of the C_2 bands; thus the results for CO replacement may be rather less reliable. However, the general effect is clear, that the light yield falls off very rapidly, about as the fourth power of the ethane concentration.

We have made similar measurements for other hydrocarbons, and results are summarized below. They are expressed as the power of the hydrocarbon concentration to which the light yield is roughly proportional. We do not necessarily attach any especial significance to a power law as such, and indeed the scatter of the experimental points and the influence of changes of temperature and pressure are such that we are not able to assign an exact form (e.g. a certain power or a particular exponential) for the curves. The power laws do, however, express the magnitude of the effect.

Results for replacement with H_2

Methane: yield proportional to rather more than square of CH_4 .

Acetylene: yield varies nearly as square of C_2H_2 .

Ethane: rather more than as square.

Propane: nearly as square, but nearer cube at very low C_2H_6 .

Benzene: slightly more rapidly than linearly.

Results for replacement with CO

Methane: light yield varies roughly as sixth power of CH_4 concentration.

Acetylene: about as cube.

Ethane: rather more than as fourth power.

Propane: rather more than as fourth power.

Benzene: about as cube.

Two further experiments were made on replacements. In a stoichiometric oxygen/acetylene flame the acetylene was progressively replaced with ethane. The curve was markedly concave downwards, but did not show a minimum, that is, the light yield dropped fairly rapidly when acetylene was replaced with ethane, more rapidly than by simply linearly combining the strength of the C_2 bands in the acetylene and ethane flames. At the other end, however, replacement of ethane with acetylene caused a slight slow rise in intensity, not a reduction.

The other observation was on replacement of acetylene with CO in a rich mixture maintained at $\lambda = 0.5$. For this the C_2 light yield again varied approximately as the cube of the acetylene concentration, or perhaps a little more rapidly.

Preliminary results for CH light yield show similar or even more striking variations. For replacement with CO the results are closely similar to those for C_2 . For H_2 replacement, however, the falling off in CH light yield tends to be more rapid than for C_2 . Thus the CH yield varies for methane nearly as the fourth power of the methane concentration and for acetylene nearly as the cube.

DISCUSSION OF RESULTS

In attempting to apply a sort of spectroscopic kinetics to flames, the interpretation of the results needs much more care than with chemical kinetics. The light yield will depend on the rate of formation and removal of the C_2 molecules, but it will also depend on any factors influencing the excitation of the C_2 , since the observations relate to the concentration of excited C_2 molecules. In the previous paper (Gaydon & Wolfhard 1948) we have reported measurements of effective rotational temperatures for C_2 and indicated that while the results cannot readily be reconciled with true chemiluminescence the excitation is certainly not of a purely thermal nature. Probably organic flames contain energy-rich molecules which give abnormally high excitation temperatures; these may be collision complexes, perhaps of a peroxidic character, retaining their heat of formation as excess vibrational energy. The strength of C_2 emission may therefore depend on the concentration of these active molecules as well as on the C_2 concentration.

The absolute light yield

For the oxygen/acetylene flame at 5.5 mm. pressure we have seen that the number of C_2 light quanta emitted per second is 6×10^{16} . Now for the Swan bands of C_2 the transition probability is probably quite high. Dr L. Brewer (personal communication) has obtained an f value around 0.5; the value could hardly exceed 1, although it might be lower. Taking the value as 0.5, the radiative life of an excited C_2 molecule would be 8×10^{-9} sec. Combining the radiative life and the number of quanta

emitted, we find that the flame contains at any instant 5×10^8 excited C_2 molecules. This number might be increased, but could not be brought lower than 2.5×10^8 (i.e. for $f = 1.0$). The volume of the reaction zone of the flame is about 5 cm^3 . Hence the concentration of excited C_2 molecules is about 1×10^8 per cm^3 , equivalent to a partial pressure of 4×10^{-11} atm. at the flame temperature.

Brewer, Gilles & Jenkins (1948) have determined the partial pressure of C_2 molecules in equilibrium with hot solid carbon. At 2700°K , which is the approximate theoretical flame temperature, their curve gives a partial pressure of 9×10^{-10} atm. This is, of course, for unexcited C_2 molecules.

For the flame at $\frac{1}{100}$ atm. pressure the theoretical maximum temperature is 2700°K , and the usual Boltzmann factor, $e^{-E/kT}$, would give a ratio of excited to unexcited molecules of 0.000035. Thus for thermal excitation of C_2 in equilibrium with solid carbon we should expect a partial pressure of excited C_2 of

$$9 \times 10^{-10} \times 0.000035 = 3 \times 10^{-14} \text{ atm.}$$

The observed partial pressure of excited C_2 is 4×10^{-11} , that is, over 1000 times greater than for equilibrium using Brewer, Gilles & Jenkins data. If we assume an abnormally high electronic excitation temperature, as is probably necessary, for the reaction zone of the flame, the discrepancy is reduced considerably, but it still seems that if excitation is due to excitation of C_2 and not to its formation in the excited state (i.e. chemiluminescence) then the concentration of C_2 molecules must be appreciably greater than that reported in equilibrium by Brewer *et al.*

At atmosphere pressure there is a modest fall in light yield. Taking the total number of excited C_2 present in the flame as the product of the light yield/ O_2 molecule entering, the number of O_2 molecules entering, and the radiative life, we obtain the figure $(3.6 \times 10^{-5}) \times (20 \times 2.69 \times 10^{19}) \times (8 \times 10^{-9}) = 1.5 \times 10^8$ for the number of excited C_2 present in the flame at any instant. The volume of the reaction zone at atmospheric pressure is very much smaller; the thickness of the zone is 0.002 cm. and the surface area about 0.1 cm^2 giving a volume of 0.0002 cm^3 . Thus at the flame temperature the partial pressure of excited C_2 is 3.4×10^{-7} atm. The theoretical flame temperature is about 3320°K , and the partial pressure of normal C_2 molecules in equilibrium with this number of excited molecules would be 1.4×10^{-3} atm. This may be compared with the figure of 5×10^{-6} atm. for the partial pressure of normal C_2 in equilibrium with solid carbon given by curves from the paper by Brewer *et al.* at this temperature. The discrepancy is less than at low pressure, but the factor of nearly 300 could not be accounted for completely by assuming a high electronic excitation temperature.

The effect of pressure

We have seen that the light yield per molecule consumed in the flame varies with pressure. For true chemiluminescence we might expect that the basic reaction producing the excited molecules would not be affected by pressure, but that if the excited molecules were quenched by collision then pressure would affect this quenching and so the light yield. At high pressure the light yield should be inversely proportional to pressure, while at low pressure the quenching effect will become less important and the light yield should approach a constant value, provided the

chemical processes are not affected by pressure. In the flame at atmospheric pressure a molecule with normal collision cross-section will make about 3×10^9 collisions per sec., so that a C_2 molecule with radiative life around 8×10^{-9} sec. might make about twenty-four collisions before radiating. This might allow some quenching. At lower pressure, however, quenching could not be important. The observations indicate a variation of light yield by a factor of 3 between 3.5 and 28 mm. pressure, and this could not be accounted for by quenching. Unless we assume that the chemical processes themselves are sensitive to pressure in this range, we must therefore reject the view that the C_2 emission results from a true chemiluminescence.

Similarly, we may reject indirect chemiluminescence in which the C_2 is activated by participation as third body in an exothermic reaction, because the number of three-body collisions will become relatively small at low pressure, and the light yield should fall instead of rising as the pressure is reduced.

If excitation is either thermal or pseudo-thermal, due to collisions with energetic molecules in the flame gases, then the strength of the radiation will depend on the strength of the excitation (which may be expressed as an effective electronic excitation temperature), and on the number of C_2 molecules available for excitation. This number will vary with the volume of the reaction zone, which is roughly inversely proportional to the cube of the pressure, and with the concentration of C_2 ; if the C_2 is formed by evaporation of carbon particles or incipient particles, then this concentration may behave like a vapour pressure and vary with flame temperature. The preceding calculations on the light yield at $\frac{1}{100}$ and 1 atm. show that for C_2 in equilibrium with solid carbon the effect of change of vapour pressure because of the different temperatures of the flames at these pressures is roughly sufficient to counter the change in the volume of the reaction zone. For flames which are cooler than the oxygen/acetylene flame, the effect of dissociation in limiting the maximum flame temperature is less important, so that the temperature will vary less with pressure; in this case the variation in vapour pressure will be less, and the volume effect will be dominant, so that the light yield will rise more markedly as the pressure is reduced; we may note that this is actually the case for propane. Thus the effect of pressure is not necessarily contrary to a theory of thermal or pseudo-thermal excitation, but lack of knowledge of the effective excitation temperatures and of the nature of carbonaceous particles in the reaction zone prevents a more detailed discussion.

Effect of fuel, mixture strength, dilution and replacement

The C_2 bands are emitted most strongly by very rich mixtures, and when the hydrocarbon is partially replaced by CO or H_2 there is a big fall in C_2 light yield. This at once suggests that the reactions leading to the formation of C_2 involve more than one hydrocarbon molecule. Certainly in the case of methane, containing only one carbon atom, there must be at least one stage of polymerization to form C_2 . This is, indeed, shown by the high power of the methane concentration to which the light yield is proportional in the replacement experiments with methane and CO. The general effects for methane and other hydrocarbons are, however, surprisingly similar, and it seems that some polymerization is necessary for all the saturated

hydrocarbons, and perhaps even for acetylene and benzene. We have pointed out that we must consider the amount of both C_2 and of the active molecules responsible for excitation. It would be possible, therefore, to explain a light yield (in the replacement experiments) depending on the square of the fuel concentration if the concentration of active species depended on the amount of fuel. However, preliminary measurements on excitation temperatures using lead lines does not seem to indicate any very marked effect of mixture strength on the excitation, and in any case even for acetylene and benzene replacement with CO gives roughly a third power law. Thus it seems that some polymerization is always necessary. Our view is that the C_2 is formed in the breaking up of some relatively complex molecules, probably of a very unsaturated character. The formation of these depends on a rapid chain polymerization process.

Partial replacement of the hydrocarbon by H_2 has a less serious effect on the light yield than an equivalent replacement with CO, and it seems that H_2 may be able to maintain the chain reaction whereas CO cannot; reaction of H atoms with hydrocarbons to form free radicals and of free radicals with H_2 to liberate H atoms are, of course, well known.

Dilution with O_2 has a very marked effect in decreasing the light yield; this may be contrasted with the effect of N_2 which is comparatively slight. The effect of O_2 may be due either to its interference with the polymerization chains or to its reaction with C_2 already formed. If the growth of the polymer was a simple chain reaction terminated by oxygen then we might be able to make deductions about the degree of polymerization from the replacement experiments with CO; this would indicate polymers of between two and six molecules in size. However, it is possible that we should regard the polymer as growing a unit on collision with a fuel molecule and usually decreasing a unit on collision with oxygen, but with the chain being unbroken. In this case the ratio of hydrocarbon to oxygen and the relative probabilities of reaction on collision would determine the likelihood of the polymer growing or disappearing, and so would not give an indication of the actual final size of the polymer, which might conceivably be quite large.

The light yield is far greater for unsaturated than for saturated hydrocarbons, and the former apparently require less polymerization. For the saturated hydrocarbons the final product may be unsaturated. It is known that in the combustion of methane it is possible to chill out acetylene (Sachse 1949). In the final stage we require a reaction which is sufficiently exothermic to liberate C_2 in its ground electronic state but with a high effective rotational temperature. If a relatively large radical were to react with O_2 , then the breaking up of the complex into CO and other fragments might liberate C_2 and enough energy. It is in any case very difficult to suggest a plausible mechanism for the formation of C_2 and liberation of energy using only simple molecules or radicals.

For saturated hydrocarbons, stripping of H atoms by reaction with radicals or free H atoms is possible, although not appreciably exothermic, but as soon as rearrangement of the molecule to form double or triple bonds occurs the C-H bonds become strengthened, and reactions involving further removal of H atoms are endothermic. Thus we cannot find any exothermic reactions with acetylene which

would liberate C_2 . Reactions of polymers such as butadiyne (biacetylene) might be more promising. The only reaction which we know of which is sufficiently exothermic to liberate C_2 and give enough energy for its electronic excitation is $C_4 + O_2 = C_2 + 2CO$, but the formation of C_4 would first have to be explained.

The relation between C_2 emission and solid carbon formation

Before entering into a discussion of this subject, we should like to stress that we are here discussing only pre-mixed flames. In these the carbon formation, when it occurs, must take place in a very short time as the gases pass through the reaction and preheating zones, and that it occurs in the presence of both fuel and oxygen. It is at least possible that the process of carbon formation may be different in diffusion flames, in which the process is comparatively slow and in which the concentration of oxygen in the heated fuel is likely to be small.

In pre-mixed flames there are three obvious possibilities:

- (i) solid carbon is formed by polymerization of C_2 ,
- (ii) C_2 is formed by evaporation of hot carbon particles or incipient particles,
- (iii) the formation of C_2 and solid carbon occur independently, perhaps by competing mechanisms.

The first possibility, that C_2 molecules polymerize to solid carbon, was suggested and favoured by Smith (1940). The present quantitative investigations, however, appear strongly against this hypothesis. The concentration of C_2 radicals in a flame cannot be very high, and the time of passage through the thin reaction zone is so short that collisions between C_2 and C_2 would be relatively rare and quite insufficient in number to account for the formation of carbon particles which are usually stated to lie between 100 and 2000 Å in size. Thus in a stoichiometric oxygen/acetylene flame at 1 atm. we have obtained a figure of 3.4×10^{-7} atm. for the partial pressure of excited C_2 molecules. In equilibrium this would correspond to a partial pressure of 1.4×10^{-3} atm. for unexcited C_2 . The time of passage through the reaction zone is only 3×10^{-7} sec. (see Gaydon & Wolfhard 1949a). For a molecule of normal collision cross-section the number of collisions suffered at atmospheric pressure and the flame temperature is about 3×10^9 per sec. Thus on the average a C_2 molecule would make about $(3 \times 10^9) \times (3 \times 10^{-7}) \times (1.4 \times 10^{-3}) = 1.4$ collisions with other C_2 molecules in passing through the luminous reaction zone. This assumes equilibrium between normal and excited C_2 molecules; actually we know that the proportion of excited molecules is much higher than for equilibrium at the flame temperature. (Wolfhard (1939) found that even with limited resolving power the blackening in the C_2 lines corresponded to a brightness temperature in excess of 2600° K, while the emissivity, deduced from failure to detect self-absorption, was less than 0.02, this indicating an excitation temperature for C_2 in excess of 4000° K.) Thus the number of normal C_2 molecules in the flame must be appreciably less than the equilibrium value, giving even fewer collisions between C_2 and C_2 . Also we must remember that collision of one C_2 with another, is not very likely to lead to polymerization at a simple bimolecular collision; a third body would probably be necessary. These calculations are for a stoichiometric mixture, which does not give

appreciable carbon formation, but with only a slightly rich mixture the acetylene flame at atmospheric pressure develops a bright white tip to the inner cone. There can be little doubt that the concentration of C_2 is quite inadequate to account for carbon formation in the short time available in premixed flames.

Considering the second possibility, it does not seem that C_2 can be formed from evaporation of actual solid particles. There is no thermal continuum emitted by flames of around stoichiometric mixtures at low pressure, as we should expect if carbon particles were present, but the C_2 emission is quite strong and our calculations show that the C_2 concentration is quite high. Indeed, if Brewer's values for C_2 vapour pressure are correct the concentration of C_2 , even after allowing for abnormal excitation, comes higher than for equilibrium with carbon for stoichiometric acetylene flames, and higher still for slightly rich flames which still do not show carbon formation at low pressure. It would also be difficult to account for the comparatively small effect of N_2 on the light yield, as the fall in flame temperature should affect the vapour pressure to an extent which could not be offset completely by the increase in volume of the reaction zone. We have also found (Gaydon & Wolfhard 1950) that the effective rotational temperature for the C_2 bands is abnormally high; this could not be expected to be the case if the C_2 were formed by normal evaporation from carbon particles.

It does, however, seem that there is a rather close parallel between carbon formation and C_2 emission. Both processes appear to depend, in pre-mixed flames, on polymerization processes which occur during the very brief time of passage through the reaction zone. Generally, the effect of pressure on C_2 emission follows a similar trend to that which we should expect for the variation of a vapour pressure in flames at different gas pressure. Also we have found experimentally that just as partial replacement with H_2 reduces the C_2 light yield from stoichiometric mixtures, so in very rich oxygen/acetylene flames it reduces the tendency to formation of solid carbon. In our present state of knowledge it seems rather likely that the C_2 emission depends on the exothermic breaking up of relatively low polymers of unsaturated character, while the formation of solid carbon requires the cracking of rather higher polymers, leading to the formation of particles of fairly large size. The smaller polymers may be regarded as incipient carbon particles exerting, owing to their possibility of exothermic decomposition, an abnormally high vapour pressure of C_2 . It is possible that the decomposition or vaporization of these smaller polymers at low pressure, where they have a longer life between collisions, may account for the reduced tendency for carbon formation at low pressure.

The emission of CH bands also seems to require polymerization, perhaps to an even higher stage than for C_2 , but the effective rotational temperature for CH is not so obviously abnormal. The presence of O_2 as diluent has much less effect on the light yield for CH, and the relation between CH emission and solid carbon formation is less close than for C_2 and solid carbon.

In conclusion, we wish to express our sincere thanks to Sir Alfred Egerton for his interest in this work, and one of us (A.G.G.) is indebted to the Royal Society for the award of a Warren Research Fellowship.

REFERENCES

- Brewer, L., Gilles, P. W. & Jenkins, F. A. 1948 *J. Chem. Phys.* **16**, 797.
 Gaydon, A. G. & Wolfhard, H. G. 1948 *Proc. Roy. Soc. A*, **194**, 169.
 Gaydon, A. G. & Wolfhard, H. G. 1949a *Proc. Roy. Soc. A*, **196**, 105.
 Gaydon, A. G. & Wolfhard, H. G. 1949b *Proc. Roy. Soc. A*, **199**, 89.
 Gaydon, A. G. & Wolfhard, H. G. 1950 *Proc. Roy. Soc. A*, **201**, 561.
 Sachse, H. 1949 *Chem. Ing. Techn.* **21**, 129.
 Smith, E. C. W. 1940 *Proc. Roy. Soc. A*, **174**, 110.
 Wolfhard, H. G. 1939 *Z. Phys.* **112**, 107.

Radially symmetric phase growth controlled by diffusion

BY F. C. FRANK, *University of Bristol*

(Communicated by N. F. Mott, F.R.S.—Received 21 November 1949)

Particular solutions of the diffusion equation, with radial symmetry, in three and in two dimensions, found originally by Rieck, represent the diffusion field around a spherical or cylindrical new phase, growing from a negligible initial radius in an initially uniform medium, maintaining equilibrium conditions at the growing surface. The resulting diffusion field is most simply described by saying that the radial gradient is the same as that of the corresponding potential problem with fixed boundaries, multiplied by a factor $\exp(-r^2/4Dt)$, where r is radius, D diffusivity, t time. The result is applied to phase growth controlled by the diffusion of heat, solute, or both together. It differs appreciably from the static approximation unless the supercooling, or the supersaturation and the solubility, are small.

1. INTRODUCTION

The diffusion problems associated with phase growth involve a number of complications, associated with their peculiar boundary conditions. (a) In general the boundary condition involves both the 'potential' and its normal gradient. (b) In the case of a crystal the boundary is commonly polyhedral; this, in association with (a), introduces serious difficulties. (c) The boundary moves, and in a manner dependent on the diffusion field.

The one-dimensional problem of an advancing flat freezing front, the rate of advance being governed by the diffusion away from the front of the latent heat of solidification, was solved by Neumann, and various workers have dealt with variants of this problem (see Carslaw & Jaeger 1947). The necessary particular solutions of the diffusion equation for spherical and cylindrical growth were found by Rieck (1924), and are quoted by Huber (1939). This work, being published only in a dissertation, has been generally overlooked, and various workers have relied on an approximate treatment which ignores the motion of the boundary (Valeton 1923, 1924; Volmer 1939; Humphreys-Owen 1949). As appears below, this approximation is permissible only when dealing with small supercooling, or the diffusion of solutes of low solubility. This is not the case with sodium chlorate, Humphreys-Owen's experimental material, and the motion of the boundary cannot then be neglected.

Rieck's work not being accessible, the problem is treated anew from the beginning in this paper. Particular solutions of the diffusion equation are found both in

spherical and cylindrical symmetry which are then shown to represent the growth of a new phase, starting from zero radius in a uniform medium, and maintaining constant 'equilibrium' conditions at the spherical surface of the growing new phase (the so-called Nernst condition). The equations may be applied equally to the diffusion of heat, or material, or both together.

In order to have equations of general applicability, the terms thermal diffusivity (thermal conductivity, as ordinarily defined, divided by density and specific heat) and material diffusivity (diffusion flux in g./sq.cm./sec. divided by concentration gradient in g./cm.³/cm.) are used. All diffusivities are measured in sq.cm./sec.⁻¹. They must be distinguished from 'diffusion constants' measured in such units as g.cm.⁻¹/sec.⁻¹ (flux in g./sq.cm./sec. divided by concentration gradient in g./100 g./cm.). In liquids, the typical order of magnitude of a thermal diffusivity is 10⁻³ sq.cm.sec.⁻¹, or more in liquid metals; of a material diffusivity 10⁻⁵ sq.cm.sec.⁻¹, or less in the more viscous liquids. In gases they are alike, both being about 0.1 to 1 sq.cm./sec.⁻¹. The diffusion equation for material cannot be valid in simple form, with diffusivities independent of concentration, unless mixing occurs without change of volume. We assume both to be true; then it follows that solvent and solute must have equal diffusivity. It is then frequently advantageous to measure concentrations by volume fraction. Various dimensionless parameters are introduced as we proceed; particularly important are the relative supercooling (actual supercooling divided by that needed for adiabatic freezing) and the analogous measure of supersaturation, actual supersaturation divided by volume fraction of solvent.

2.1. SOLUTION OF THE DIFFUSION EQUATION FOR SPHERICAL SYMMETRY

Let us consider the growth, with spherical symmetry, of a new phase, controlled by the diffusion of a single entity which may be either heat (the latent heat of formation of the new phase) or matter (a solvent, or impurity). In all its applications this involves some approximation, justified to a greater or lesser extent. Convection has to be neglected which is seldom completely justified save in very viscous or rigid media, when additional complications due to mechanical stress arise. We may be using the sphere to represent approximately a compact polyhedral crystal. Sometimes the crystal may be truly spherical (e.g. spherulitic); but in that case the spherical form is not necessarily stable; there will probably be a tendency for it to develop dendritic branches. The geometrical conditions apply more strictly to a droplet growing either from vapour or from a second liquid. Nevertheless, the case considered corresponds sufficiently closely to many important cases in practice to be the most important which we can treat theoretically.*

* It is in fact possible to extend the treatment employed below to the case of a crystal growing concentrically with itself in *any constant shape*, and maintaining constant conditions at corresponding points on its surface, the result being obtained in the form of a series of spherical harmonics multiplied by functions of the reduced radius defined below. The first function in the series, with spherical symmetry, corresponds precisely to the case discussed in this paper. All the higher functions have a mean value of zero on the surface, and zero net flux through the surface of any sphere about the origin, and approach zero with increase of radius more rapidly than does the first. Thus the result discussed below suffices to calculate the growth rate of such a crystal, and also the form of the diffusion field in its outer regions, or anywhere if averaged with respect to angular co-ordinates.

We wish to see whether there can be any quasi-stationary solution of the diffusion equation which has spherical symmetry. For dimensional reasons, if there is any such solution, it must be one in which the diffusion potential (i.e. temperature, or concentration) is a function of the dimensionless 'reduced radius'

$$s = rD^{-1/2}t^{-1/2}, \quad (1)$$

where r is the radius, D the diffusivity and t the time,

$$\phi(r, t) = \phi(s). \quad (2)$$

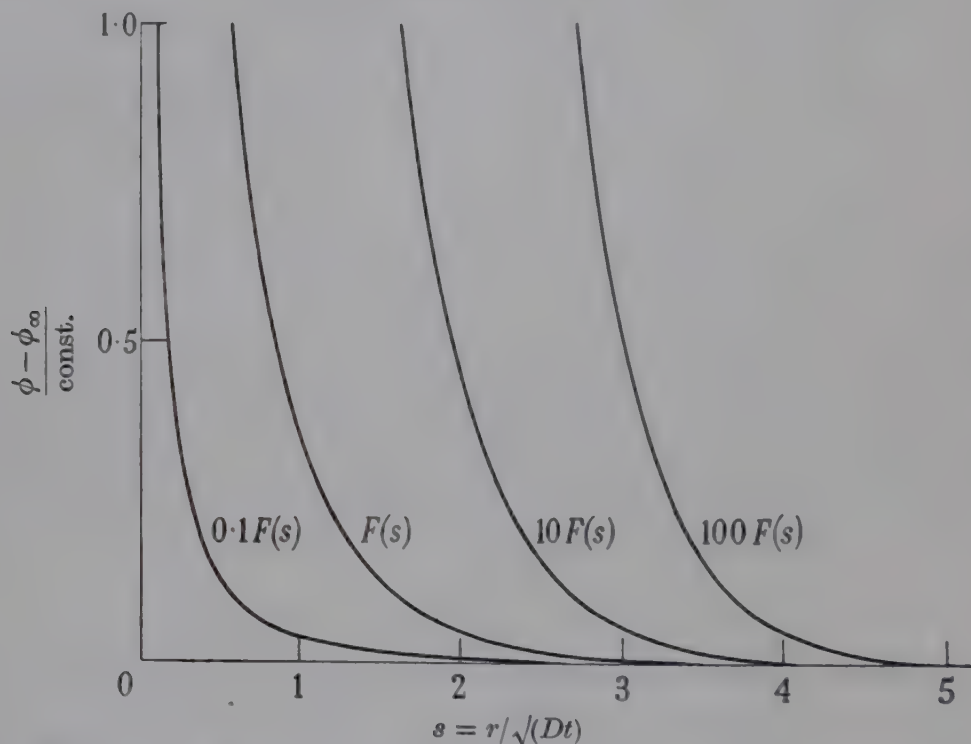


FIGURE 1. The growth diffusion function for spherical symmetry:

$$F(s) = s^{-1} \exp(-\frac{1}{4}s^2) - \frac{1}{2}\pi^{1/2}[1 - \operatorname{erf}(\frac{1}{2}s)].$$

Though the curves merely represent one and the same function on various scales, they may be interpreted as showing the temperature (or reduction in concentration) as a function of radius around a new phase growing spherically so that its radius R is proportional to $t^{1/2}$, when $S = R/(Dt)^{1/2}$ is 0.099, 0.58, 1.61 and 2.73 respectively; thus showing the transition from an approximately reciprocal law when the radius is small to an approximately exponential law when it is large.

Then

$$\left. \begin{aligned} \frac{\partial \phi}{\partial t} &= \frac{\partial \phi}{\partial s} \frac{\partial s}{\partial t} = -\frac{s}{2t} \frac{\partial \phi}{\partial s}, \\ \frac{\partial \phi}{\partial r} &= \frac{\partial \phi}{\partial s} \frac{\partial s}{\partial r} = D^{-1/2} t^{-1/2} \frac{\partial \phi}{\partial s}, \\ \frac{\partial^2 \phi}{\partial r^2} &= D^{-1} t^{-1} \frac{\partial^2 \phi}{\partial s^2}. \end{aligned} \right\} \quad (3)$$

Thus the diffusion equation

$$\frac{\partial \phi}{\partial t} = D \nabla^2 \phi = D \left(\frac{\partial^2 \phi}{\partial r^2} + \frac{2}{r} \frac{\partial \phi}{\partial r} \right) \quad (4)$$

(terms for angular dependence being omitted) becomes

$$-\frac{s}{2t} \frac{\partial \phi}{\partial s} = \frac{1}{t} \left(\frac{\partial^2 \phi}{\partial s^2} + \frac{2}{s} \frac{\partial \phi}{\partial s} \right),$$

i.e.

$$\frac{\partial^2 \phi}{\partial s^2} = - \left(\frac{s}{2} + \frac{2}{s} \right) \frac{\partial \phi}{\partial s}. \quad (5)$$

TABLE 1. THE GROWTH-DIFFUSION AND DIFFUSION-GROWTH FUNCTIONS

s or S	spherical case 3 dimensions		cylindrical case 2 dimensions	
	$F(s)$	$f(S)$	$F_2(s)$	$f_2(S)$
0	∞	0	∞	0
0.1	9.1388	0.4580×10^{-2}	2.7059	0.01356
0.2	4.1639	0.01682	2.0190	0.04080
0.3	2.5176	0.03476	1.5973	0.07351
0.4	1.7133	0.05706	1.3406	0.1116
0.6	0.9282	0.1097	0.9594	0.1890
0.8	0.5586	0.1678	0.7046	0.2646
1.0	0.3539	0.2272	0.5221	0.3352
1.2	0.2303	0.2852	0.3872	0.3996
1.4	0.1521	0.3405	0.2860	0.4576
1.6	0.1010	0.3923	0.2098	0.5094
1.8	0.06716	0.4402	0.1525	0.5554
2.0	0.04454	0.4843	0.0097	0.5964
2.2	0.02938	0.5245	0.0780	0.6327
2.4	0.01924	0.5612	0.0547	0.6650
2.6	0.01249	0.5946	0.0379	0.6937
2.8	0.008020	0.6250	0.0258	0.7193
3.0	0.005095	0.6525	0.0174	0.7421
3.2	0.003197	0.6776	0.0115	0.7625
3.4	0.001981	0.7004	0.00751	0.7808
3.6	0.001200	0.7211	0.00483	0.7972
3.8	0.0007296	0.7400	0.00304	0.8120
4	0.0004334	0.7572	0.00190	0.8254
5	0.00002544	0.8236	0.000150	0.8757
6	$10^{-6} \times 0.9910$	0.8673	0.000006	0.9078

Putting $\partial \phi / \partial s = p$ we have $\ln p = -\frac{1}{4}s^2 - 2 \ln s + A_1$,

$$\partial \phi / \partial s = -As^{-2} \exp(-\frac{1}{4}s^2). \quad (6)$$

$$\text{Hence } \phi = A \int_s^\infty z^{-2} \exp(-\frac{1}{4}z^2) dz + \phi_\infty, \quad (7)$$

where ϕ_∞ is the value of ϕ at infinity. Integrating by parts

$$\begin{aligned} \phi - \phi_\infty &= A \left[z^{-1} e^{-\frac{1}{4}z^2} - \int_{\frac{1}{2}s}^\infty e^{-z^2} dz \right] \\ &= A \left[s^{-1} e^{-\frac{1}{4}s^2} - \frac{1}{2} \pi^{\frac{1}{2}} \{1 - \operatorname{erf}(\frac{1}{2}s)\} \right] \\ &\equiv AF(s), \quad \text{say.} \end{aligned} \quad (8)$$

The function $F(s)$ is shown graphically in figure 1 and numerically in table 1.

Now the total flux of the diffusing entity (heat, or matter) over a sphere of radius $R \equiv SD^{\frac{1}{2}}t^{\frac{1}{2}}$ is

$$\begin{aligned} -4\pi R^2 D(\partial\phi/\partial r)_{r=R} &= -4\pi S^2 D^{\frac{1}{2}}t^{\frac{1}{2}}(\partial\phi/\partial s)_{s=S} \\ &= 4\pi A D^{\frac{1}{2}}t^{\frac{1}{2}} \exp(-\tfrac{1}{4}S^2). \end{aligned} \quad (9)$$

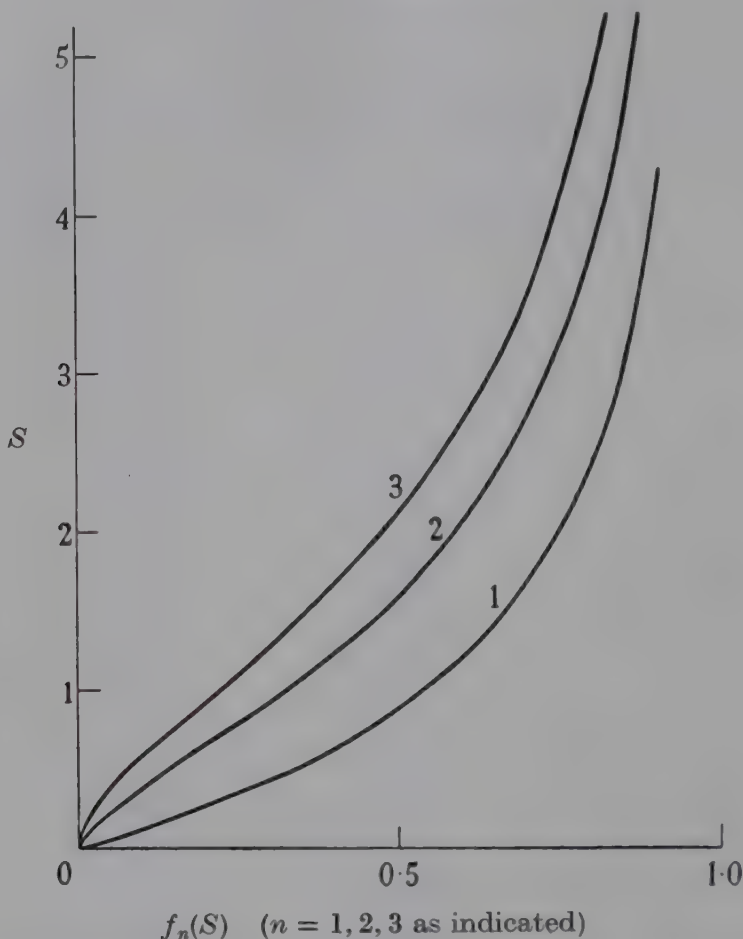


FIGURE 2. The diffusion growth functions for spherical, cylindrical, and linear cases (suffixes 3, 2, 1 respectively). S is plotted as a function of

$$f_3(S) = \tfrac{1}{2}S^3 \exp(\tfrac{1}{4}S^2) F(S),$$

$$f_2(S) = \tfrac{1}{2}S^2 \exp(\tfrac{1}{4}S^2) F_2(S),$$

and

$$f_1(S) = \tfrac{1}{2}S \exp(\tfrac{1}{4}S^2) F_1(S).$$

Note that

$$F_1(s) = \pi^{\frac{1}{2}}[1 - \operatorname{erf}(\tfrac{1}{2}s)].$$

These curves show the 'reduced radius' $r/\sqrt{(Dt)}$ (or correspondingly reduced thickness in the linear case) when growth is controlled by diffusion of heat or material, and $f(S)$ is the 'reduced supercooling' or 'reduced supersaturation' defined in the text.

Some concern must here be had for dimensions and units; if ϕ measures temperature, which we can regard as a proportional measure of heat density, the expression just obtained must be multiplied by ρc , where ρ is density and c specific heat, to represent a heat flux in the usual units.

Now the diffusing entity is expelled at the boundary of the growing phase in proportion to the amount of the latter formed (this applies either to latent heat or to solvent which is expelled from the volume claimed by the growing new phase; approximately, in any case, accurately if ϕ is constant at the surface). Let q be

the amount expelled per unit volume of new phase formed, measured in such units that ϕ measures a 'q-density' (e.g. if l is the latent heat per gram, $l\rho$ would be the value of q in heat units, but l/c is the appropriate measure for conformity with ϕ as a temperature); then, if R is now taken to represent the radius of the sphere of new phase we may equate the diffusion flux to the rate of expulsion at the surface:

$$4\pi AD^{\frac{1}{2}}t^{\frac{1}{2}}\exp(-\frac{1}{4}S^2) = q \frac{d}{dt}(\frac{4}{3}\pi R^3) \\ = q 2\pi S^3 D^{\frac{1}{2}}t^{\frac{1}{2}},$$

$$\text{i.e.} \quad A = \frac{1}{2}qS^3 \exp(\frac{1}{4}S^2). \quad (10)$$

On the surface of the growing sphere of reduced radius S , ϕ has the constant value

$$\phi_S = \frac{1}{2}qS^3 \exp(\frac{1}{4}S^2) F(S) + \phi_{\infty} \\ = qf(S) + \phi_{\infty}, \quad \text{say.} \quad (11)$$

S is shown as a function of $(\phi_s - \phi_{\infty})/q$, in accordance with this equation, in curve 3, figure 2. Numerical values of $f(S)$ are given in table 1.

It is now clear that we have obtained a solution of the diffusion equation, exact according to the declared assumptions, which corresponds to the sphere of new phase growing at a decreasing linear rate such that the radius is proportional to the square root of the time (equation (9), with S a constant determined by (8) and (11), or figure 2), and maintaining a constant equilibrium condition at its surface (for ϕ_s is constant—the only value of ϕ which would remain constant at the boundary while the rate of deposition varied is the equilibrium value). Thus this corresponds to Nernst's boundary condition. It is not possible for this to be exactly true from the very beginning when the rate of deposition per unit area is theoretically infinite; but for reasonably fast-growing phases this should usually introduce errors only at the beginning of the growth process, which may soon be allowed for by a mere shift of the time-zero, and later even totally neglected. The nature of the solution is most clearly indicated by equation (6), which shows that the gradient of ϕ is the same as in an ordinary static potential problem, multiplied by a Gaussian cut-off factor with characteristic radius $2D^{\frac{1}{2}}t^{\frac{1}{2}}$.

2.2. SERIES EXPANSIONS

$F(s)$ (equation (8)) may be expressed as a series, particularly convenient for small values of s :

$$F(s) = s^{-1} - \frac{1}{2}\pi^{\frac{1}{2}} + \frac{1}{2}(\frac{1}{2}s) - \frac{1}{1!3.4}(\frac{1}{2}s)^3 + \frac{1}{2!5.6}(\frac{1}{2}s)^5 - \dots \quad (12)$$

For large values of s , it is preferable to use the asymptotic expansion (in which the error is less than the last term used)

$$F(s) \approx 2s^{-3}e^{-\frac{1}{4}s^2} \left[1 - \frac{1.3}{2} \left(\frac{2}{s} \right)^2 + \frac{1.3.5}{2} \left(\frac{2}{s} \right)^4 - \dots \right]. \quad (13)$$

By the same expansions (11) becomes

$$\begin{aligned}
 (\phi - \phi_\infty)/q &\equiv f(S) \\
 &= 2(\tfrac{1}{2}S)^2 + 4(\tfrac{1}{2}S)^4 + \tfrac{8}{3}(\tfrac{1}{2}S)^6 + \tfrac{16}{15}(\tfrac{1}{2}S)^8 + \dots \\
 &\quad - \pi^2[2(\tfrac{1}{2}S)^3 + 2(\tfrac{1}{2}S)^5 + (\tfrac{1}{2}S)^7 + \tfrac{1}{3}(\tfrac{1}{2}S)^9 + \dots], \quad (14)
 \end{aligned}$$

and

$$f(S) \approx 1 - \frac{1.3}{2} \left(\frac{2}{S}\right)^2 + \frac{1.3.5}{2^2} \left(\frac{2}{S}\right)^4 - \dots \quad (15)$$

Thus, $(\phi - \phi_\infty)/q$ is proportional to S^2 for small S and tends asymptotically to 1 as S goes to infinity.

2.3. GROWTH CONTROLLED BY THE DIFFUSION OF HEAT OR MATERIAL

When we are dealing with the diffusion of heat, the case $(\phi_s - \phi_\infty)/q = 1$ signifies that the supercooling of the distant medium is just equal to that required for adiabatic freezing, l/c . Then (if no other process intervened to limit the growth rate) the crystal would instantaneously grow to an infinite size. For any value of $(\phi_s - \phi_\infty)/q$ larger than 1 the Nernst boundary condition cannot be satisfied; the supercooling being more than l/c , the temperature at freezing cannot rise as high as the melting-point.

If, on the other hand, we are dealing with the diffusion of material, and can disregard thermal effects, q represents the volume fraction of solvent or impurity (whichever we choose to call it) in the old phase at the surface of the new one, in the ideal case that the solvent or impurity is totally expelled from the new phase (ideal crystallization). Then $q = \phi_s$. Usually a certain amount of solvent or impurity will enter into the new phase, either in solution, or occluded in cracks for example. Let the amount so entering be $\epsilon\phi_s$, where ϵ will usually be a small fraction. Then

$$\frac{\phi_s - \phi_\infty}{q} = \frac{\phi_s - \phi_\infty}{(1 - \epsilon)\phi_s} = \frac{\psi_\infty - \psi_s}{(1 - \psi_s)(1 - \epsilon)}, \quad (16)$$

where ψ (suffix S at the surface of the new phase, ∞ at a remote point in the old phase) represents the volume fraction of solute. Thus $(\phi_s - \phi_\infty)q$ is equal to the product of three factors:

- (a) the supersaturation of the remote solution,
- (b) a factor equal to 1 when the solubility is very small, and tending to infinity when the solubility is very large, and
- (c) a factor which is unity when no impurity or solvent enters into the new phase, but larger when this occurs.

If ϵ is zero, $(\phi_s - \phi_\infty)/q$ cannot exceed unity.

2.4. GROWTH CONTROLLED BY THE DIFFUSION OF HEAT AND MATERIAL

In practice, growth is liable to be controlled by the diffusion of both heat and matter. Diffusion of matter may be the predominantly controlling factor in phase growth from a solution, but the temperature rise resulting from latent heat liberated may be sufficient to change the solubility significantly. Alternatively, as in the solidification of a melt, the diffusion of heat may be the predominantly controlling

factor, but there may be a sufficient amount of impurity present to accumulate at the surface of the solid and cause a significant depression of the melting-point. These two cases are not really distinct. A small amount of a second component is called an impurity, a large amount a solvent. We can treat as one the whole range from the extreme of a pure melt where heat diffusion is the sole controlling factor to that of a dilute solution where the diffusion of material matters alone. The solution of the diffusion equation obtained above still applies, for both diffusing entities, if we may neglect the effect that each has on the diffusion of the other. It is usually tolerable to neglect the change of thermal diffusivity with composition. It is less generally justified to neglect the dependence of a material diffusivity upon temperature. However, since the thermal diffusivity is usually a good deal larger than that of material, the important part of the material diffusion field usually lies well within the thermal diffusion field, in a region where the temperature may reasonably be treated as uniform. Hence, treating the two diffusivities as independent constants, we let θ and ϕ denote temperature and volume fraction of second component (an impurity, let us call it) and D_θ and D_ϕ their diffusivities, respectively, and put $S_\theta = RD_\theta^{-\frac{1}{2}}t^{-\frac{1}{2}}$, $S_\phi = RD_\phi^{-\frac{1}{2}}t^{-\frac{1}{2}}$. If the solid grows with radius proportional to $t^{\frac{1}{2}}$, θ and ϕ will have constant values at the interface, given by equation (11). We have to find that growth rate which makes these values θ_S and ϕ_S together correspond to the Nernst boundary condition. Thus we have to solve simultaneously:

$$\theta_S - \theta_\infty = (l/c)f(S_\theta), \quad (17.1)$$

$$\phi_S - \phi_\infty = \phi_S f(S_\phi), \quad (17.2)$$

$$\theta_S = \theta_m - K\phi_S, \quad (17.3)$$

$$s_\theta D_\theta^{\frac{1}{2}} = S_\phi D_\phi^{\frac{1}{2}}. \quad (17.4)$$

Equations (17.1) and (17.2) correspond to equation (11) and $f(S)$ is the function there defined. The effect of a non-zero incorporation factor ϵ will be considered later. Equation (17.3) expresses the depression of the melting-point from its value θ_m in pure solute by the volume fraction of impurity at the crystal surface, so that K is a depression constant which can be calculated thermodynamically. It is, in fact, equal to $RT^2/l\rho V$, where R is the gas constant, T the absolute temperature, l the latent heat per gram and ρ the density (molten) of the solute, and V the molecular volume of the impurity. As we shall see below, Kc/l determines the sensitivity of the system to impurity, so that this varies inversely as the square of the latent heat. Now, eliminating ϕ_S and θ_S from (17), we have

$$f(S_\theta) = \frac{c}{l} \left[\theta_m - \theta_\infty - \frac{K\phi_\infty}{1 - f(S_\phi)} \right]. \quad (18)$$

This equation has the same interpretation as equation (11) (with ϕ there interpreted as temperature), with the difference that the effective value of the supercooling is diminished by the depression of the freezing-point caused by the impurity accumulated at the growing surface. In liquids S_ϕ will usually be of the order 10 or more times S_θ , so that $f(S_\phi)$ may approach 1 for quite a moderate value of S_θ . Thus a small amount of impurity may have a large effect on the growth rate.

Equation (18) may be solved by successive approximations. If we prepare a graph of the functions $f(S)$ and $-\log [1-f(S)]$ against $\log S$, we may obtain the solution with the aid of an instrument, made of sliding scales of which two are logarithmic and one is linear (standard logarithmic graph papers will provide the scales). The mode of operation is shown in figure 3. When three initial settings have

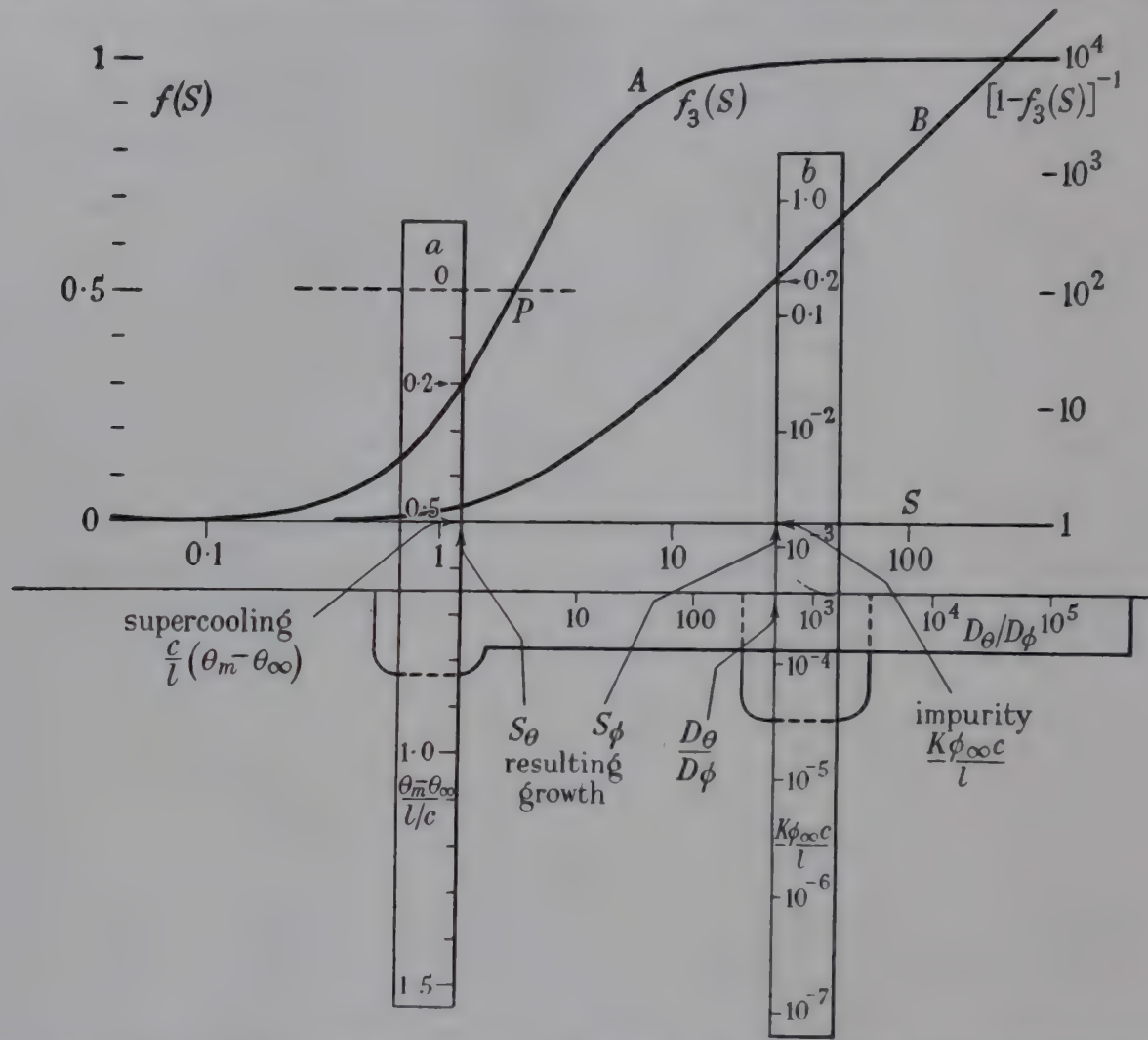


FIGURE 3. Method of solving equation (18) for growth of a spherical crystal controlled by diffusion of both heat and material. The settings shown indicate that when the reduced supercooling is 0.5 (so that if diffusion of heat alone controlled the growth rate, the reduced radius S_θ would be 2.09) the presence of an amount of impurity sufficient to depress the freezing-point by $1.6 \times 10^{-3} l/c.$, with a diffusivity 430 times smaller than the thermal diffusivity, will reduce S_θ to 1.26. The actual depression of the temperature at the surface of the crystal, below its melting-point, due to impurity accumulated at the surface is then $0.2 l/c.$

been made, the instrument is slid horizontally until the reading of curve A on scale a is the same as the reading of curve B on scale b. This solves the problem:

$$\frac{c}{l}(\theta_m - \theta_\infty) - f(S_\theta) = \frac{K\phi_\infty c}{l} \frac{1}{1 - f(s_\theta D_\theta^\dagger / D_\phi^\dagger)}.$$

If $K\phi_\infty c/l$ is very small (very little impurity present) the value of S_θ is determined by the crossing point P. With increasing amounts of impurity it is increasingly depressed; ultimately A becomes practically constant and the diffusion of matter determines the growth rate almost exclusively.

A special case of some interest is that in which the supercooling just suffices for adiabatic freezing, i.e.

$$\theta_m - \theta_\infty = l/c.$$

Then (18) becomes $[1 - f(S_\theta)][1 - f(S_\phi)] = K\phi_\infty c/l.$ (19)

If now the amount of impurity is very small, both $f(S_\theta)$ and $f(S_\phi)$ may be replaced by the leading terms of the asymptotic expansion (15), giving

$$S_\theta^2 S_\phi^2 = 36l/cK\phi_\infty,$$

i.e. $S_\theta = (36D_\phi/cK\phi_\infty D_\theta)^{\frac{1}{2}}.$ (20)

For example, if l/c is 100°C , $K\phi_\infty$ is 0.01°C , and $D_\theta/D_\phi = 100$, then $S_\theta = 7.8$, $S_\phi = 78$. If the absolute value of the thermal diffusivity is then 0.001 sq.cm./sec. , we have $R = 0.78t^{\frac{1}{2}}\text{ cm.}$

2.5. EFFECT OF INCORPORATION OF IMPURITY INTO THE NEW PHASE

If an amount of impurity equal to $\epsilon\phi_s$ is incorporated into the new phase, the right-hand side of (17.2) must be multiplied by $(1 - \epsilon)$. Supposing this incorporation occurs by the occlusion of saturated solution, from which no latent heat is liberated, the right-hand side of (17.1) should be multiplied by $(1 - \epsilon)$ also; but this is of trivial effect, and may be disregarded. We obtain

$$f(S_\theta) = \frac{c}{l} \left[\theta_m - \theta_\infty - \frac{K\phi_\infty}{1 - (1 - \epsilon)f(S_\phi)} \right]. \quad (21)$$

Now, in contrast with the previous case, the impurity cannot be raised in concentration at the growing surface, as compared with its concentration in the outer medium, by a factor of more than ϵ^{-1} . If it is raised by such a factor, the amount of impurity incorporated into the new phase will, incidentally, be equal to that in the outer medium.

Equation (21) may be solved by the graphical method described above, if we prepare a similar graph having curves of

$$-\log [1 - (1 - \epsilon)f(S_\phi)].$$

In the special case of supercooling sufficient for adiabatic freezing, we have

$$[1 - f(S_\theta)][1 - (1 - \epsilon)f(S_\phi)] = K\phi_\infty c/l. \quad (22)$$

If the amount of impurity is very small, so that S_θ and S_ϕ are both large, we have, approximately,

$$S_\theta = (6\epsilon l/cK\phi_\infty)^{\frac{1}{2}}. \quad (23)$$

2.6. GROWTH FROM SOLUTION

The notation of equation (17) et seq., wherein we work essentially with temperatures (expressing, in effect, the amount of impurity ϕ by the depression of freezing-point, $K\phi$, it would cause), is appropriate to the discussion of freezing influenced by a small amount of impurity. For solutions, however, it is more suitable to deal with the volume fraction of solute, $\psi = 1 - \phi$. We shall let ψ_∞^* represent the solubility of the solute (measured as a volume fraction) at the temperature θ_∞ . Thus

$$\psi_\infty^* = 1 - (\theta_m - \theta_\infty)/K. \quad (24)$$

This strictly assumes a linear dependence of volume fraction solubility on temperature from the melting-point down, but it suffices if this relation is linear in the actual working range, since we shall eliminate θ_m and can define K from the temperature coefficient of solubility:

$$1/K = d\psi^*/d\theta. \quad (25)$$

(18) now becomes
$$f(S_\theta) = \frac{Kc}{l} \left[1 - \psi_\infty^* - \frac{1 - \psi_\infty}{1 - f(S_\psi)} \right]. \quad (26)$$

S_ψ is, of course, the same as S_ϕ . This equation is of the same form as (18) and may be solved in the same way. For some purposes we may prefer it in the rearranged form

$$1 - f(S_\psi) = \frac{1 - \psi_\infty}{1 - \psi_\infty^* - (l/cK) f(S_\theta)}. \quad (27)$$

When the solubility is small, $f(S_\theta)$ may safely be approximated to $\frac{1}{2}S_\theta^2$, i.e. $\frac{1}{2}S_\psi^2 D_\psi/D_\theta$. D_ψ is the same as D_ϕ . We now have

$$f(S_\psi) \approx \psi_\infty - \psi_\infty^* - \frac{1}{2}(l/cK) S_\psi^2 D_\psi/D_\theta. \quad (28)$$

If it is permissible to approximate $f(S_\psi)$ similarly, we have

$$\frac{1}{2}S_\psi^2 [1 + (l/cK) D_\psi/D_\theta] = \psi_\infty - \psi_\infty^*, \quad (29)$$

i.e. apart from a thermal correction, which may amount to a few per cent, the reduced radius S_ψ is given by the square root of twice the supersaturation expressed as a volume fraction. For values of S_ψ larger than permit the approximation of (28) to (29), the thermal correction is somewhat, though only moderately, greater.

2.7. GROWTH OF A COMPOUND

If two dissolved substances which diffuse independently† unite to form the new phase, in such a way that equilibrium is maintained at the spherical growing surface, our solution of the diffusion equation can again be applied. We shall consider that thermal effects, and incorporation of solvent in the new phase, are both negligible. Let the concentrations of the two components be $\phi^{(1)}$ and $\phi^{(2)}$ ‘parts’ per unit volume, where ‘parts’ may, at choice, be grams, moles, or millilitres in the liquid state, since additivity of volumes is an essential approximation. Let unit volume of the new phase contain $\alpha^{(1)}$ and $\alpha^{(2)}$ ‘parts’ of the two components. Then the equations which apply simultaneously are:

$$\phi_\infty^{(1)} - \phi_S^{(1)} = (\alpha^{(1)} - \phi_S^{(1)}) f(S^{(1)}), \quad (30.1)$$

$$\phi_\infty^{(2)} - \phi_S^{(2)} = (\alpha^{(2)} - \phi_S^{(2)}) f(S^{(2)}), \quad (30.2)$$

$$\phi_S^{(1)} \phi_S^{(2)} = C, \quad (30.3)$$

$$S^{(1)}/S^{(2)} = (D^{(2)}/D^{(1)})^{\frac{1}{2}}. \quad (30.4)$$

† This assumption of independent diffusion is unlikely to be true unless the concentrations of both solutes are small, or their diffusivities equal. If in a three-component mixture the diffusivities of two components are independent of concentration, and unequal, that of the third cannot be independent of concentration. But when all three are present in comparable amounts, the departure from simple behaviour is unlikely to fall on one alone.

(30.3) is the 'solubility product' relationship. Elimination of $\phi_S^{(1)}$ and $\phi_S^{(2)}$ gives

$$\frac{[\phi_\infty^{(1)}/\alpha^{(1)} - f(S^{(1)})][\phi_\infty^{(2)}/\alpha^{(2)} - f(S^{(2)})]}{[1 - f(S^{(1)})][1 - f(S^{(2)})]} = \frac{C}{\alpha^{(1)}\alpha^{(2)}}. \quad (31)$$

This equation may be solved in essentially the same way as equation (18) by use of a graph of $f(S)$ against $\log S$.

3.1. GROWTH CONTROLLED BY TWO-DIMENSIONAL DIFFUSION

It is experimentally difficult to determine the three-dimensional concentration or temperature field around a growing particle, especially as the scale of the experiment must be microscopic in order that convection shall be negligible. Bunn (1949), Berg (1938) and Humphreys-Owen (1949) have observed the concentration around a growing crystal sandwiched between flat plates. The cells they used were scarcely thin enough for total neglect of convection, but the worst errors from this source occur near the surface of the crystal, where the gradient is steepest. The theory here developed is in any case inapplicable in this region because of the shape of the crystal boundary, and in the outer regions of the diffusion field, where a fair correspondence is to be expected, convection should introduce smaller errors. In any case, it should be possible to perform similar experiments with thinner cells, and make convection genuinely negligible (Frank 1949). When the diffusion problem is thus reduced to a two-dimensional one, we have in place of equation (4)

$$\frac{\partial \phi}{\partial t} = D \left(\frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} \right), \quad (32)$$

$$\text{in place of (5)} \quad \frac{\partial^2 \phi}{\partial s^2} = - \left(\frac{s}{2} + \frac{1}{s} \frac{\partial \phi}{\partial s} \right), \quad (33)$$

$$\text{in place of (6)} \quad \partial \phi / \partial s = -As^{-1} \exp(-\frac{1}{4}s^2), \quad (34)$$

$$\begin{aligned} \text{and in place of (8)} \quad \phi - \phi_\infty &= \frac{1}{2}A \int_{\frac{1}{4}s^2}^{\infty} z^{-1} e^{-z} dz \\ &= -\frac{1}{2}A \cdot \text{Ei}(-\frac{1}{4}s^2) \equiv AF_2(s), \quad \text{say.} \end{aligned} \quad (35)$$

Ei is the exponential integral, tabulated, for example, by Jahnke & Emde (1943). The function $F_2(s)$ is shown graphically in figure 4, and numerically in table 1.

Now the total flux of the diffusing entity over a cylinder of radius $R = SD^{\frac{1}{2}}t^{\frac{1}{2}}$ and height h is

$$\begin{aligned} -2\pi RhD(\partial \phi / \partial r)_{r=R} &= -2\pi ShD(\partial \phi / \partial s)_{s=S} \\ &= 2\pi hDA \exp(-\frac{1}{4}S^2). \end{aligned}$$

If q signifies, once again, the amount of the diffusing entity expelled on the formation of unit volume of the new phase, we may equate the diffusion flux to the rate of expulsion at the surface,

$$\begin{aligned} 2\pi hDA \exp(-\frac{1}{4}S^2) &= q(d/dt)(\pi R^2 h) \\ &= q\pi S^2 Dh, \end{aligned}$$

$$\text{i.e.} \quad A = \frac{1}{2}qS^2 \exp(\frac{1}{4}S^2). \quad (36)$$

Hence on the surface of the growing cylinder of reduced radius S , ϕ has the constant value

$$\phi_S = -\frac{1}{4}qS^2 \exp(\frac{1}{4}S^2) \text{Ei}(-\frac{1}{4}S^2) + \phi_\infty \equiv qf_2(S) + \phi_\infty, \quad \text{say.} \quad (37)$$

The function $f_2(S)$ is shown numerically in table 1 and graphically in figure 4. The further analysis of two-dimensional problems can proceed in precise analogy to the three-dimensional problems treated above, the functions $F(s)$ and $f(S)$ being replaced by $F_2(s)$ and $f_2(S)$. In the case of Bunn's, Berg's and Humphreys-Owen's experiments, however, we do not meet an analogous situation with respect to the simultaneous diffusion of heat and matter, because heat can escape through the confining walls whereas matter cannot; it is easy to assure oneself, however, that thermal effects can be neglected in these experiments. The nature of the particular

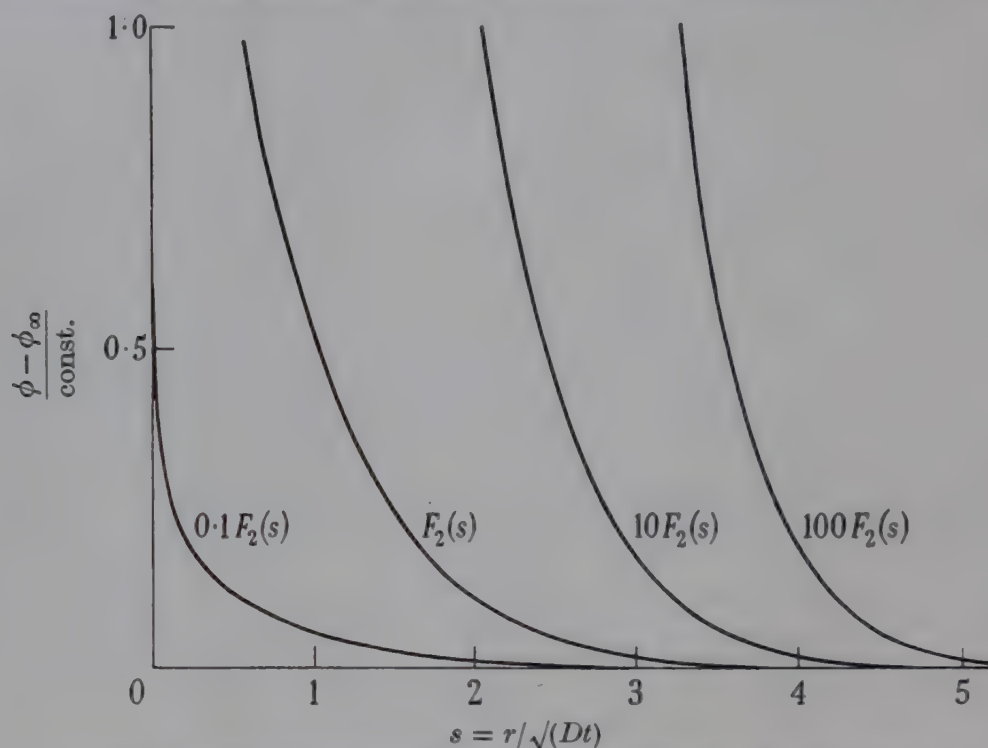


FIGURE 4. The growth-diffusion function for cylindrical symmetry:

$$F_2(s) = -\frac{1}{2} \text{Ei}(-\frac{1}{4}s^2).$$

The curves may be interpreted as graphs of temperature or of reduction in concentration, as a function of radius around a new phase growing cylindrically so that its radius R is proportional to $t^{\frac{1}{2}}$, when $S = R/(Dt)^{\frac{1}{2}}$ is 6×10^{-5} , 0.58, 2.06 and 3.28 respectively. The law is approximately exponential for the larger radii. For the smallest we have a transition from the logarithmic law in the innermost part to the exponential law in the outermost part of the diffusion field.

solution obtained is most clearly seen from equation (32). As in the three-dimensional case, the radial gradient is the same as in the corresponding static potential problem, multiplied by a Gaussian cut-off factor, which is the same as in the three-dimensional case.

3.2. EXPANSIONS AND APPROXIMATIONS

For small values of S we may use the expansion

$$\begin{aligned} F_2(s) &\equiv -\frac{1}{2} \text{Ei}(-\frac{1}{4}s^2) \\ &= \frac{1}{2} \ln(4/\gamma s^2) - \frac{1}{2}(\frac{1}{2}s)^2 + \frac{1}{2 \cdot 2 \cdot 2!} (\frac{1}{2}s)^4 - \frac{1}{2 \cdot 3 \cdot 3!} (\frac{1}{2}s)^6 + \dots \\ &= 0.404539 - \ln s - \frac{1}{2}(\frac{1}{2}s)^2 + \frac{1}{2 \cdot 2 \cdot 2!} (\frac{1}{2}s)^4 - \dots \\ &= \ln(1.4984/s) - \frac{1}{2}(\frac{1}{2}s)^2 + \frac{1}{2 \cdot 2 \cdot 2!} (\frac{1}{2}s)^4 - \dots \end{aligned} \quad (38)$$

Thus for small values of the reduced radius, s , the diffusion potential $\phi - \phi_\infty$ is well represented by the logarithmic formula which would be anticipated from approximating the problem by static potential theory, and extrapolates to zero at $S = \frac{3}{2}$, very nearly. Actually, $F_2(s)$ is then 0.245. By the same expansion

$$\begin{aligned} f_2(S) &\equiv 2(\tfrac{1}{2}S)^2 \exp(\tfrac{1}{2}S)^2 F_2(S) \\ &= 2(\tfrac{1}{2}S)^2 \exp(\tfrac{1}{2}S)^2 \ln(1.4984/S) - (\tfrac{1}{2}S)^4 - \tfrac{3}{4}(\tfrac{1}{2}S)^6 - \tfrac{11}{36}(\tfrac{1}{2}S)^8 - \tfrac{25}{288}(\tfrac{1}{2}S)^{10} - \dots \end{aligned} \quad (39)$$

For very large values of s we have

$$F_2(s) \approx 2s^{-2} \exp(-\tfrac{1}{4}s^2),$$

and accordingly

$$f_2(S) \approx 1.$$

My thanks are due to Miss M. J. Littleton for performing the computations for table 1, and to Mr G. Wyllie and Mr R. G. Wylie for suggestions concerning the graphical-instrumental solution of equation (18).

REFERENCES

- Berg, W. F. 1938 *Proc. Roy. Soc. A*, **164**, 79.
 Bunn, C. W. 1949 *Disc. Faraday Soc.* no. 5 (Crystal Growth), p. 132.
 Carslaw, H. S. & Jaeger, J. C. 1947 *Conduction of heat in solids*, pp. 71, 227. Oxford: University Press.
 Frank, F. C. 1949 *Disc. Faraday Soc.* no. 5 (Crystal Growth), p. 189.
 Humphreys-Owen, S. P. F. 1949 *Proc. Roy. Soc. A*, **197**, 218.
 Jahnke, E. & Emde, F. 1943 *Tables of functions* (Dover reprint), p. 6.
 Rieck, R. 1924 See Huber, 1939, *Z. angew. Math. Mech.* **19**, 1.
 Valetton, J. J. P. 1923 *Z. Krystallogr.* **59**, 168, 335.
 Valetton, J. J. P. 1924 *Z. Krystallogr.* **60**, 1.
 Volmer, M. 1939 *Kinetik der Phasenbildung*, p. 209. Leipzig: Steinkopff.

The far ultra-violet absorption spectra of the hydrides and deuterides of sulphur, selenium and tellurium and of the methyl derivatives of hydrogen sulphide

By W. C. PRICE,* *Wheatstone Laboratory, King's College, London*
AND J. P. TEEGAN AND A. D. WALSH, *Department of Chemistry, University of Leeds*

(Communicated by J. T. Randall, F.R.S.—Received 25 November 1949)

[Plates 15 and 16]

The absorption spectra in the vacuum ultra-violet of the hydrides and deuterides of sulphur, selenium and tellurium, and methyl mercaptan and dimethyl sulphide are described. Well-developed Rydberg series leading to the following ionization potentials have been found: H_2S , 10.47 V; MeSH , 9.44 V; H_2Se , 9.88 V; H_2Te , 9.14 V. In the case of one series for H_2Se fifteen members of the series were observed. The spectra of the deuterides are almost identical with those of the hydrides, showing that virtually every band in the spectra is due to a separate electronic transition. This and the general nature of the rotational fine structure show the transitions concerned to be those of an electron from a non-bonding ground-state orbital, i.e. from the p lone-pair ground-state orbital. The nature of the upper orbitals of the various series is also interpreted and shown to provide explanations of certain peculiarities of the observations. The quantity $I(X) - I(\text{H}_2X)$, where X is a group VI element, or $I(Y) - I(\text{HY})$, where Y is a group VII element, is shown to be positive and comparatively large when X or Y lies in the first period of the periodic table, but to change sign and to remain almost constant at a small negative value as one passes to elements in later periods. A plot of $I(\text{H}_2X)$ against the first ionization potential of the corresponding inert gas is linear. Extrapolation enables the first ionization potential of H_2Po to be predicted at 8.6 V. A similar plot for the halogen acids, if assumed linear, yields a predicted first ionization potential for HF of 17.0 ± 0.7 V.

INTRODUCTION

The absorption spectra of H_2O and H_2S in the vacuum ultra-violet have previously been described (Price 1936*a*). The present paper extends that work by describing and discussing the absorption spectra of H_2Se and H_2Te . It provides information on the spectroscopic changes, and associated changes in electronic structure, that accompany the change of an atom in a molecule for an atom of higher atomic number but in the same group of the periodic table. The only extensive earlier work on this theme is that on the halogen acids and their alkyl derivatives (Price 1936*b*, 1938). In order to facilitate the interpretation of the spectra of H_2S , H_2Se and H_2Te , photographs of the corresponding spectra of D_2S , D_2Se and D_2Te have been obtained and are described here. Quantitative information on the effect of alkyl substitution upon the spectrum and electronic structure of H_2S is provided by an account of the far ultra-violet absorption spectra of MeSH and Me_2S .

A preliminary brief note upon the present work has already been published (Price 1948). All the spectra were obtained with a 1 m. spectrograph, using a diffraction grating of 30,000 lines to the inch ruled on glass. The dispersion was about 8 Å/mm. As all the bands reported here appear at extremely low pressures ($c.$ 0.01 mm.

* The early experimental work described in this paper was carried out at the Ryerson Physical Laboratory, University of Chicago.

in path lengths of 2 m.), only small quantities of the absorbing gases were needed. These were conveniently prepared in a special apparatus by dropping water on to an intimate mixture of phosphorus pentoxide with ferrous sulphide, ferrous selenide or aluminium telluride respectively. For the preparation of the deuterides 99 % D_2O was used.

THE SPECTRA OF H_2S AND D_2S

Figure 1*a* and *b*, plate 15, shows the absorption spectra of H_2S and D_2S in the wave-length range 2000 to 1100 Å. The spectrum of H_2S has already been described (Price 1936*a*) but is included here for comparison with the other spectra reported.

H_2S has a region of weak, continuous, absorption between 3000 and 2000 Å, part of which can just be distinguished on the extreme left of the photographs in figure 1. The first discrete band appears at about 1580 Å, and thenceforward to shorter wave-lengths the spectrum reveals a large number of discrete bands. Most of these have prominent rotational structure, consisting of strong and sharp *Q* branches accompanied on either side by weaker *R* and *P* branches. The heads of these *Q* branches fit the following Rydberg series formulae:

$$\nu_0^n = 84,420 - \frac{R}{(n - 2.04)^2} \quad (n = 5, 6, 7, \dots, 14), \quad (1)$$

$$\nu_0^n = 84,520 - \frac{R}{(n - 1.57)^2} \quad (n = 4, 5, 6, \dots, 10). \quad (2)$$

The limit of series (1) corresponds to the value 10.466 ± 0.002 V. The limits of the two series are not quite identical (cf. methyl acetylene, Price & Walsh 1945); the mean of the two yields a value 10.47 V for the first ionization potential of H_2S .

The bands of series (2) and (1) were respectively referred to as the *A* and *E* bands in the original paper (Price 1936*a*). The former have *P* branches that are extremely weak compared with their *R* branches. The latter have *P* and *R* branches of similar intensity. To the long wave-length side of all the *E* bands there appear strong, diffuse bands called the *D* bands. To the long wave-length side of the *D* bands another set of bands called *C* is found. These are sharper than the *D* bands and are similar to those of the *A* system in that they possess *R* branches much stronger than the *P* branches. Both the *C* and *D* bands approach the *E* bands as *n* increases in series (1). This classification covers all the bands observed below 1460 Å. It regards them all as due to purely electronic transitions. Weaker bands in the region from 1600 to 1460 Å were explained as a set of vibration bands accompanying the main electronic transitions in that region.

The extreme sharpness of the *Q* branches and the fact that the rotational structure in the *P* and *R* branches is evenly spaced show that the change in moment of inertia in passing from the ground to the excited states concerned in series (1) and (2) is very slight. This conclusion is supported by the fact that according to the above analysis nearly all the bands are due to pure electronic excitation without any accompanying vibration. This is what might be expected for the excitation of an electron from a non-bonding orbital.

Figure 1*b* shows the spectrum of D₂S. The bands show practically no shift relative to H₂S. This implies that every band is a different electronic transition and so places the interpretation of the spectrum of H₂S on a firmer footing. Even the weaker bands in the region 1600 to 1460 Å (e.g. at 1579 and 1545 Å and between 1516 and 1460 Å) are identical in position to within the accuracy of measurement, i.e. to within 10 cm.⁻¹. If, as the earlier analysis supposed, they were vibrational bands, a frequency diminution on passing from H₂S to D₂S in the ratio $\approx 1/\sqrt{2}$ would be expected. This would mean a shift of at least several hundred wave numbers for a band such as that at 1545 Å which is separated from the electronic transition at 1516 Å by 1250 cm.⁻¹. It is clear therefore that the interpretation of these bands must be revised; like those at shorter wave-lengths they are due to vibrationless electronic transitions. This conclusion reinforces the above arguments that the spectrum is due to the excitation of an electron from a non-bonding orbital.

THE SPECTRA OF METHYL MERCAPTAN AND DIMETHYL SULPHIDE

The spectrum of methyl mercaptan is shown in figure 1*c*. At the longest wave-lengths (2000 to 1800 Å) continuous absorption is visible. This is part of the continuous absorption below 2780 Å that has been reported previously (Hukomoto 1936). A diffuse band occurs at *c.* 1840 Å followed by a system of discrete bands in the region 1760 to 1700 Å. Another system begins at *c.* 1640 Å. There is little doubt that these two systems correspond to the H₂S bands in the neighbourhood of the first members of series (2) and (1) respectively. They are shifted some 200 Å to the red as a result of the methyl substitution. Their rotational fine structure, distinct in H₂S, becomes blurred in MeSH. At shorter wave-lengths a large number of discrete bands appear. It is fairly easy to pick out corresponding bands of H₂S and MeSH and so to find the following Rydberg series for the latter:

$$\nu_0^n = 76138 - \frac{R}{(n-2.05)^2} \quad (n = 5, 6, 7, \dots, 13), \quad (3)$$

$$\nu_0^n = 76197 - \frac{R}{(n-1.55)^2} \quad (n = 4, 5, 6, \dots, 9). \quad (4)$$

Series (3) corresponds to series (1) (the *E* bands) and series (4) to series (2) (the *A* bands) of H₂S. Tables 1 and 2 show the agreement between observed and calculated frequencies.

TABLE 1. TABLE SHOWING THE OBSERVED AND CALCULATED FREQUENCIES OF SERIES (3)

<i>n</i>	$\nu_{\text{obs.}}$ (cm. ⁻¹)	$\nu_{\text{calc.}}$ (cm. ⁻¹)	Δ (cm. ⁻¹)
5	—	63528	—
6	69124	69105	+ 19
7	71660	71659	+ 1
8	73038	73038	0
9	73882	73866	+ 16
10	obscured (?)	74402	—
11	74750	74768	- 18
12	75025	75029	- 4
13	75217	75223	- 6

From considerations of intensity and general position, the $n = 5$ member of series (3) lies within, or is represented by, the 1640 to 1580 Å system. The strongest member of this system has a wave number of $c. 61800 \text{ cm.}^{-1}$. This is a long way from the value calculated from series (3) for $n = 5$; but marked disturbance of such a low-lying Rydberg band is to be expected. The $n = 4$ member of series (4) is similarly expected to lie within, or be represented by, the 1760 to 1700 Å region; the strongest band of the system lies at $c. 57250 \text{ cm.}^{-1}$.

TABLE 2. TABLE SHOWING THE OBSERVED AND CALCULATED FREQUENCIES OF SERIES (4)

n	$\nu_{\text{obs.}}$ (cm.^{-1})	$\nu_{\text{calc.}}$ (cm.^{-1})	Δ (cm.^{-1})
4	—	57715	—
5	66968	66977	− 9
6	70650	70655	− 5
7	72506	72502	+ 4
8	obscured (?)	73559	—
9	74220	74220	0

Some of the members of a third Rydberg series can also be picked out. The series formula is

$$\nu_0^n = 76170 - \frac{R}{(n - 1.30)^2} \quad (n = 4, 5, 6, \dots, 9). \quad (5)$$

Table 3 shows the observed and calculated frequencies. The bands of this series approach those of series (3) as one passes to shorter wave-lengths. Probably the bands correspond to the *C* and *D* bands of H_2S .

TABLE 3. TABLE SHOWING THE OBSERVED AND CALCULATED FREQUENCIES OF SERIES (5)

n	ν_{obs} (cm.^{-1})	$\nu_{\text{calc.}}$ (cm.^{-1})	Δ (cm.^{-1})
4	61144	61117	+ 27
5	68174	68154	+ 20
6	71185	71202	− 17
7	72807	72792	+ 15
8	73729	73725	+ 4
9	74315	74319	− 4

The limits of series (3), (4) and (5) are nearly identical. That of series (3) corresponds to $9.437 \pm 0.002 \text{ V}$. The extrapolation to the limit is only over $c. 900 \text{ cm.}^{-1}$.

Like the spectra of H_2S and MeSH , the spectrum of Me_2S shows weak continuous absorption between 3000 and 2000 Å. At $c. 2025 \text{ Å}$ (see figure 1*d*) a diffuse region of absorption appears which is probably analogous to the MeSH region at $c. 1840 \text{ Å}$. It is followed by a system of strong bands stretching from $c. 1960$ to $c. 1850 \text{ Å}$. This presumably corresponds in part to the MeSH 1760 to 1700 Å system and perhaps also to the MeSH 1640 to 1580 Å system. Frequency differences of $c. 650 \text{ cm.}^{-1}$ and $c. 330 \text{ cm.}^{-1}$ are noticeable in this region. Weaker, diffuse absorption appears at

c. 1795 Å. At 1645 Å strong and fairly sharp bands begin which no doubt correspond to the main Rydberg series members of H₂S and MeSH. Unfortunately, comparatively few of these could be measured, since they rapidly merge into continuous absorption in the region 1530 to 1400 Å. However, if, as seems probable, the 1645 Å band corresponds to that at 1493 Å in MeSH a drop of ≈ 0.7 V in first ionization potential from MeSH to Me₂S is indicated.

TABLE 4. TABLE SHOWING THE OBSERVED AND CALCULATED FREQUENCIES OF SERIES (6)

<i>n</i>	ν_{obs} (cm. ⁻¹)	$\nu_{\text{calc.}}$ (cm. ⁻¹)	Δ (cm. ⁻¹)
6	67098	67093	+ 5
7	72704	72670	+ 34
8	75228	75224	+ 4
9	76608	76603	+ 5
10	77433	77431	+ 2
11	77966	77967	- 1
12	78338	78333	+ 5
13	78604	78597	+ 7
14	78786	78790	- 4
15	78938	78936	+ 2
16	79047	79050	- 3
17	79140	79140	0
18	79217	79213	+ 4
19	79273	79273	0
20	79320	79322	- 2

Photographs of the spectra of ethyl mercaptan and diethyl sulphide in the region 2250 to 1700 Å have been published previously (Price 1935). The first strong absorption of EtSH occurs at *c.* 2050 Å and that of Et₂S in the region 2250 to 2200 Å. The first absorption of Et₂S is thus shifted to long wave-lengths relative to EtSH by an amount very similar to the corresponding shift in Me₂S relative to MeSH.

THE SPECTRA OF H₂Se AND D₂Se

The spectrum of H₂Se is shown in figure 2*a*, plate 16. A region of diffuse absorption has its maximum at *c.* 1970 Å. Strong discrete bands begin at 1690 Å and can be seen particularly well to crowd to a limit around 1255 Å. A prominent Rydberg series of formula

$$\nu_0^n = 79703 - \frac{R}{(n - 3.05)^2} \quad (n = 6, 7, 8, \dots, 20) \quad (6)$$

can be picked out. Table 4 shows the agreement between observed and calculated frequencies.

No less than fifteen members of series (6) could be observed. In only one case, namely, acetaldehyde (Walsh 1946), has a greater number of members of a Rydberg series for a polyatomic molecule been observed. The extrapolation to the limit is over less than 400 cm.⁻¹ and yields a value for the first ionization potential of H₂Se of 9.879 ± 0.001 V.

Less well-developed Rydberg series can also be distinguished, leading to approximately the same limit. One such series has the formula

$$\nu_0^n = 79677 - \frac{R}{(n - 2.15)^2} \quad (n = 5, 6, 7, \dots, 9). \quad (7)$$

Table 5 shows the agreement between observed and calculated frequencies for series (7). The early members of series (6) and (7) have strong and sharp *Q* branches placed symmetrically between fairly strong *P* and *R* branches. The bands of series (6) and (7) correspond respectively to the *D* and *E* bands of H_2S . The *C* bands of H_2S also have obvious analogues in the spectrum of H_2Se .

TABLE 5. TABLE SHOWING THE OBSERVED AND CALCULATED FREQUENCIES OF SERIES (7)

<i>n</i>	$\nu_{\text{obs.}}$ (cm.^{-1})	$\nu_{\text{calc.}}$ (cm.^{-1})	Δ (cm.^{-1})
5	66179	66167	+ 12
6	72262	72274	- 12
7	75025	75012	+ 13
8	76485	76470	+ 15
9	77325	77338	- 13

Another probable series leading to roughly the same limit as (6) and (7) begins with the strong band at $1650 \text{ \AA}$ and has a Rydberg denominator of $(n - 2.55)^2$. Only about four members can be traced however. The members correspond to the *A* bands of H_2Se . With both H_2O and H_2S the Rydberg series starting with the longer wave-length strong bands are weaker and less well developed than those starting with the bands at shorter wave-lengths. The same observation applies with even more force to H_2Se . The long series (6) has as its first discrete member the comparatively short wave-length band at $1491 \text{ \AA}$.

Figure 2*b* shows the spectrum of D_2Se . As with D_2S relative to H_2S , it shows practically no change relative to H_2Se . It is clear that virtually every band in the spectrum of H_2Se is to be interpreted as a separate electronic transition. The ionization potential of D_2Se , like that of D_2S relative to H_2S , is not appreciably different from that of H_2Se .

THE SPECTRA OF H_2Te AND D_2Te

The spectrum of H_2Te is shown in figure 2*c*. Diffuse absorption appears at *c.* $2000 \text{ \AA}$, a weak but discrete band at $1916 \text{ \AA}$ and many strong, discrete bands to short wave-lengths of $1880 \text{ \AA}$. The spectrum has an obvious general resemblance to those of H_2S and H_2Se . There are repeated patterns of bands converging to a limit near $1360 \text{ \AA}$. A well-developed Rydberg series can again be picked out, starting with the strong band at $1689 \text{ \AA}$ and having the formula

$$\nu_0^n = 73705 - \frac{R}{(n - 3.95)^2} \quad (n = 7, 8, 9, \dots, 14). \quad (8)$$

Table 6 shows the agreement between observed and calculated frequencies.

Other series of bands can be distinguished, some of which are marked in figure 2*c*. None of these, however, can be traced very far to short wave-lengths. As with H_2S and H_2Se , it is noticeable that the most extensive Rydberg series does not begin with the longest wave-length discrete bands. The $n = 7$ member of (8) gives only a poor fit in the series formula, but the general appearance of the series seems definitely to indicate the $1689\text{\AA}$ band as the first discrete member. The limit of the series corresponds to a value $9.138 \pm 0.005\text{ V}$ for the first ionization potential of H_2Te .

TABLE 6. TABLE SHOWING THE AGREEMENT BETWEEN OBSERVED AND CALCULATED FREQUENCIES OF SERIES (8)

n	$\nu_{\text{obs.}}$ (cm.^{-1})	$\nu_{\text{calc.}}$ (cm.^{-1})	Δ (cm.^{-1})
7	61559	61908	- 449
8	67016	67015	+ 1
9	69408	69402	+ 6
10	70703	70707	- 4
11	71497	71497	0
12	72012	72012	0
13	72350	72365	- 15
14	72621	72619	+ 2

As with H_2S and H_2Se , the rotational fine structure of many of the bands is prominent. The three strong bands in the region 1880 to $1820\text{\AA}$ all have sharp Q branches and strongly developed R branches, as have also their analogues at shorter wave-lengths marked in figure 2. These correspond to the A system of H_2S . The strong band at $1621\text{\AA}$ and its analogues at shorter wave-lengths (also marked in figure 2) have both P and R branches prominent on either side of the line-like central Q branch. These and the members of series (8) correspond to the C, D, E system of H_2S .

Figure 2*d* shows the spectrum of D_2Te . It is practically identical with the spectrum of H_2Te . The ionization potential of D_2Te is not appreciably different from that of H_2Te .

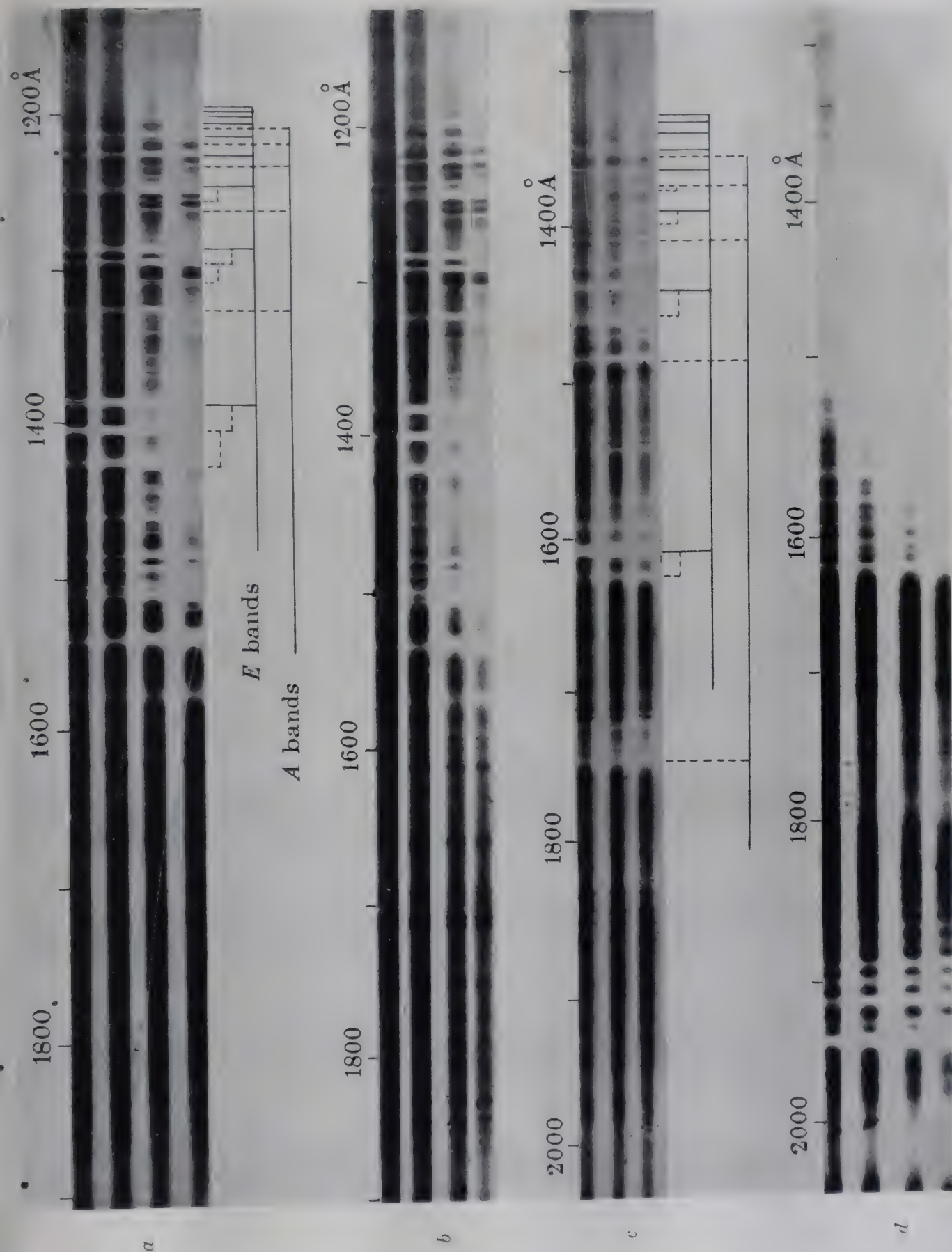
GENERAL DISCUSSION

The absence of vibrational structure (shown by comparison of the hydrides and deuterides), the nature of the rotational fine structure and the general correspondence of all the spectra reported here make it certain that the spectra are due to the excitation of an electron from a non-bonding orbital. The fact that the spectra are, in terms of energy, the most easily excited shows further that the orbital concerned is that of the lone-pair p electrons on the S, Se or Te atoms.

In the case of H_2O , the following Rydberg series has been found (Price 1936):

$$\nu_0^n = 101,780 - \frac{R}{(n - 1.05)^2} \quad (n = 4, 5, 6, \dots, 10). \quad (9)$$

The limit corresponds to 12.61 V . The structure of the bands and the general appearance of the spectra show that series (1), (3), (6), (8) and (9) all correspond. These are the series that in each case are the most well developed.

FIGURE 1. The absorption spectra of *a*, H_2S ; *b*, D_2S ; *c*, MeSH ; and *d*, Me_2S .

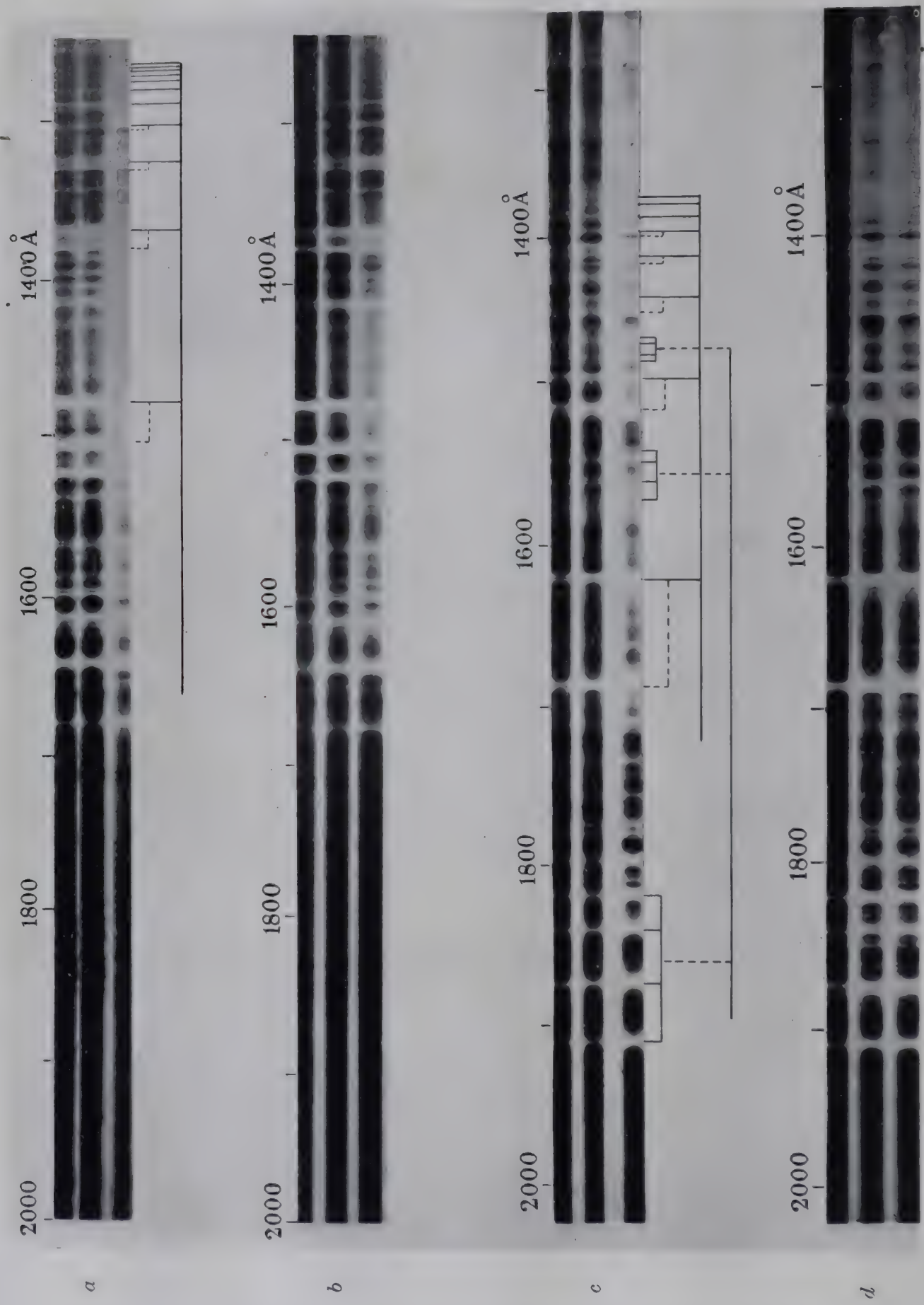


FIGURE 2. The absorption spectra of *a*, H₂Se; *b*, D₂Se; *c*, H₂Te; and *d*, D₂Te.

The conclusion that the series are due to the excitation of an electron from a lone-pair p orbital, i.e. from an orbital in the molecules closely similar to the corresponding atomic orbital in the separate group VI element, leads one to look for correlations between the series and the atomic Rydberg series known for O, S, Se and Te. The closeness of the molecular ionization potentials to the corresponding ionization potentials of the atoms (table 7) shows how little the p lone-pair orbital is affected by the change from X to H_2X . Values of atomic ionization potentials used in this paper are taken from Herzberg (1937) but increased by 0.5 % to allow for the change since 1937 in the most probable value of the electronic charge. Table 7 shows how, although the quantity $I(X) - I(H_2X)$ is positive and appreciable when X is oxygen, it changes sign on passing to sulphur and then remains almost constant at a small negative value on passing to selenium and tellurium. Similar behaviour is shown in table 8 for the halogen acids. The ionization potentials for the halogen acids are taken from Price (1936*b*) (a correction being applied for the change in the value of the electronic charge) with the exception of that for HF which is derived below; the figures represent means of the limits of the series leading to the doublet states $^2\Pi_{\frac{1}{2}, \frac{1}{2}}$ of the molecular ion. The figures of table 8 are subject to greater uncertainty than those of table 7, but they show a value for $I(X) - I(HX)$ which is again probably appreciable and positive when X is of low atomic weight but becomes small and negative as one passes to the group VII elements of higher atomic weight.

TABLE 7

H_2X	$I(X)$ (V)	$I(H_2X)$ (V)	Δ (V)
H_2O	13.62	12.61	+1.01
H_2S	10.35	10.47	-0.12
H_2Se	9.75	9.88	-0.13
H_2Te	9.00	9.14	-0.14

TABLE 8

HX	$I(X)$ (V)	$I(HX)$ (V)	Δ (V)
HF	17.43	17.00 ± 0.7	$+0.4 \pm 0.7$
HCl	13.01	$12.8(9) \pm 0.1$	$+0.1 \pm 0.1$
HBr	11.86	$12.0(9) \pm 0.1$	-0.2 ± 0.1
HI	10.6	$10.7(6) \pm 0.1$	-0.2 ± 0.1

Price (1936*a*) showed that the term values of series (9) for H_2O were closely related to the $2p^3(^4S)ns$ terms of O(I). Similarly, the series (1) for H_2S , (3) for $MeSH$, (6) for H_2Se , (8) for H_2Te are all to be interpreted as due to transitions of an electron from a p ground-state orbital to s upper levels. The quantum defects in these series have all been chosen to bring out the resemblance to the related atomic series. It is remarkable, however, that these defects themselves (for the molecular, as for the related atomic, series) appear to be quantized, being all close to a value $m + 0.05$, where m is a small whole number. In each case an earlier member of the series is expected to lie in the region where long wave-length continuous absorption is observed. In the case of H_2O and H_2S this continuum has in fact been interpreted

by Mulliken (1935) as due to a transition of a p lone-pair electron to an s upper level.* Taking therefore the continua as the first members of the series, ' n ' begins at 3 for H_2O , 4 for H_2S and MeSH , 5 for H_2Se and 6 for H_2Te —which are the figures one would expect from the principal quantum numbers of the valence shells concerned.

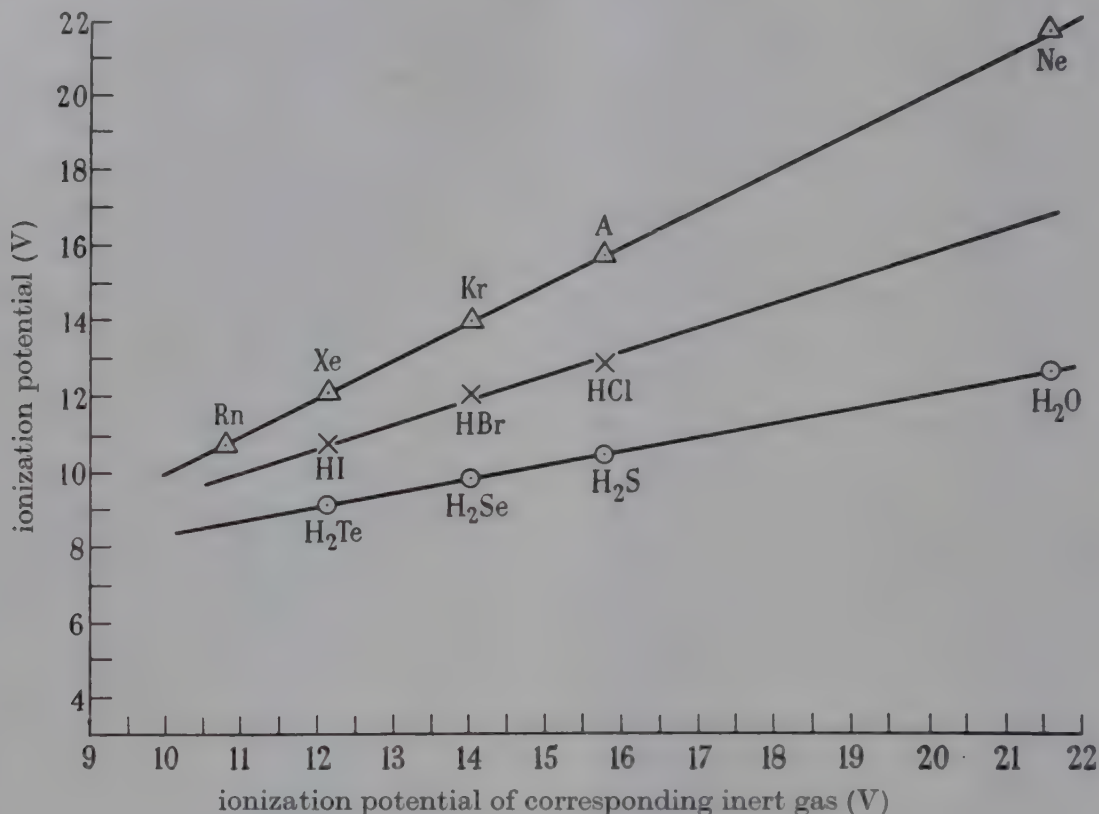


FIGURE 3

The above series are the best developed in the spectra. The series that have been found to begin with the longer wave-length discrete bands are less well developed. For H_2O and H_2S , Price showed that they almost certainly corresponded to the atomic series $2p^3(^4\text{S})np$. There is little doubt that the same applies to MeSH , H_2Se and H_2Te . That the series are due to transitions of an electron from a p ground-state orbital to a p upper-state orbital explains (1) their failure to be well developed, since the higher the excitation the more the upper orbital resembles an atomic orbital and transitions from p to p orbitals in an atom are forbidden; (2) the fact that their

* The interpretation of the long-wave continuum as a transition involving an upper orbital which is atomic in character is perhaps a little difficult to understand. However, it must be remembered that the upper orbital of this transition on account of its small energy still lies largely within the framework of the molecule and is therefore profoundly influenced by the molecular potential field. For example, it will overlay the orbitals of the HX bonds and interact strongly with them giving rise to the photodissociation effects associated with the continua. As the quantum number increases, progressively smaller fractions of the excited orbital lie within the molecular framework, and the orbitals tend to assume more completely the atomic character. The bands sharpen up because the interaction with bonding orbitals which depends on the amount of overlap also becomes progressively smaller. Thus while the Rydberg series predict a low member this could not be expected to have the same character as the upper members. As there is no region of absorption except the continuum which could be assigned to the lower member it seems that the interpretation given by Mulliken is the correct one.

quantum defects are smaller than those of the series interpreted as due to transitions to s upper levels—for in the atomic case it is known that Rydberg series with s upper levels have greater quantum defects than those with p upper levels.

The molecules H_2O , H_2S , ..., are respectively isoelectronic with the molecules HF , HCl , ..., and with the inert gases Ne , Ar , One expects, therefore, smooth relations between, for example, the p lone-pair ionization potentials of these three series. Figure 3 shows in fact that a plot of the first ionization potentials of the hydrides of group VI against the first ionization potentials of the corresponding inert gases is not only smooth but linear. A short extrapolation to the ionization potential of Rn enables the first ionization potential of H_2Po to be predicted at 8.6 V . Figure 3 also shows a plot of the first ionization potentials of the hydrides of I , Br and Cl against the ionization potentials of the corresponding inert gases. On the assumption that this plot is linear, the first ionization potential for HF is predicted at $17.0 \pm 0.7 \text{ V}$. The linearity of the plot is supported not only by the analogy with the graph for H_2O , H_2S , ..., but also by the fact that it lies close to half-way between the lines for the inert gases and for F_2 , S_2 , Clearly, figure 3 shows the existence of regular and mutually consistent relationships of importance for future theoretical consideration.

One of us (W.C.P.) wishes to acknowledge assistance by the U.S. Office of Naval Research under Task Order IX of Contract N6-ori 20 with the University of Chicago. Another of us (J.P.T.) wishes to thank the Commissioners for the Exhibition of 1851 for the award of an Overseas Scholarship (for Eire), during the tenure of which this work was carried out. We are also grateful to the Chemical Society for a grant for equipment used in this work.

REFERENCES

- Herzberg, G. 1937 *Atomspectra and atomic structure*. New York: Prentice Hall.
 Hukomoto, H. 1936 *Sci. Rep. Tôhoku Imp. Univ.* **25**, 1162. Quoted in Spenser, H. & Teller, E. 1941 *Rev. mod. Phys.* **13**, 75.
 Mulliken, R. S. 1935 *J. m. Phys.* **3**, 506.
 Price, W. C. 1935 *J. Chem. Phys.* **3**, 256.
 Price, W. C. 1936a *J. Chem. Phys.* **4**, 147.
 Price, W. C. 1936b *J. Chem. Phys.* **4**, 539.
 Price, W. C. 1938 *Proc. Roy. Soc. A*, **167**, 216.
 Price, W. C. 1948 *J. Chem. Phys.* **16**, 551.
 Price, W. C. & Walsh, A. 1945 *Trans. Faraday Soc.* **41**, 381.
 Walsh, A. D. 1946 *Proc. Roy. Soc. A*, **185**, 176.

INDEX TO VOLUME 201 (A)

- Adsorption of vapours on mercury. IV. Surface potentials and chemisorption (Kemball), 377.
- Andrade, E. N. da C. Wilkins Lecture, Robert Hooke, 439.
- Atmospheric waves (cellular) in the ionosphere and troposphere (Martyn), 216.
- Atomic explosion of 1945 (Taylor), 175.
- Aust, K. T. & Chalmers, B. The specific energy of crystal boundaries in tin, 210.
- Bagguley, D. M. S. & Griffiths, J. H. E. Paramagnetic resonance in copper sulphate, 366.
- Bardwell, J. & Hinshelwood, Sir Cyril. The slow oxidation of gaseous methyl ethyl ketone, 26.
- Batchelor, G. K. On the spontaneous magnetic field in a conducting liquid in turbulent motion, 405.
- Behaviour of supersonic flow past a body of revolution, far from the axis (Whitham), 89.
- Belcher, H. & Sugden, T. M. Studies on the ionization produced by metallic salts in flames. I. The determination of the collision frequency of electrons in coal-gas/air flames, 480.
- Beryllium atom, calculation for the ground state (Boys), 125.
- Blast wave, formation of, by a very intense explosion (Taylor), 159, 175.
- Boyd, R. L. F. The collection of positive ions by a probe in an electrical discharge, 329.
- Boys, S. F. Electronic wave functions. II. A calculation for the ground state of the beryllium atom, 125.
- Bremner, J. G. M. The thermal accommodation coefficients of gases. I. An investigation of the effect of flashing, 305.
- Bremner, J. G. M. The thermal accommodation coefficients of gases. II. Determinations at room and at liquid-oxygen temperatures, 321.
- C₂ bands, measurements of light yield (Gaydon & Wolfhard), 570.
- C₂ bands, rotational temperatures and excitation mechanism (Gaydon & Wolfhard), 561.
- Cellular atmospheric waves in the ionosphere and troposphere (Martyn), 216.
- Chalmers, B. *See* Aust & Chalmers.
- Childs, E. C. & Collis-George, N. The permeability of porous materials, 392.
- Chirgwin, B. H. & Coulson, C. A. The electronic structure of conjugated systems. VI, 196.
- Clark, G. L. Note on the problem of a rotating mass of perfect fluid in relativity mechanics, 510.
- Clark, G. L. The external gravitational and electromagnetic fields of rotating bodies, 488.
- Collection of positive ions by a probe in an electrical discharge (Boyd), 329.
- Collis-George, N. *See* Childs & Collis-George.
- Conductivity of thin wires (Dingle), 545.
- Conjugated systems, electronic structure (Chirgwin & Coulson), 196.
- Contact between a solid and an electrolyte (Grimley), 40.
- Copper sulphate, paramagnetic resonance (Bagguley & Griffiths), 366.
- Coulson, C. A. *See* Chirgwin & Coulson.
- Coxeter, H. S. M. & Whitrow, G. J. World-structure and non-Euclidean honeycombs, 417.
- Crystal boundaries in tin, specific energy (Aust & Chalmers), 210.
- Crystal structure of horse met-myoglobin. I. General features: the arrangement of the polypeptide chains (Kendrew), 62.
- de Wet, J. S. The interaction representation in the quantum theory of fields, 284.
- Diffraction gratings, reproduction and ruling (Merton), 187.
- Dingle, R. B. The electrical conductivity of thin wires, 545.
- Dislocations, one-dimensional. IV. (Frank & van der Merwe), 261.
- Dispersion of a pressure pulse in the atmosphere (Seorer), 137.

- Electrical conductivity of thin wires (Dingle), 545.
- Electrical discharge, collection of positive ions by a probe (Boyd), 329.
- Electrolyte, contact between a solid and an (Grimley), 40.
- Electronic structure of conjugated systems (Chirgwin & Coulson), 196.
- Electronic wave functions. II. A calculation for the ground state of the beryllium atom (Boys), 125.
- Explosion, formation of a blast wave by a very intense. I and II (Taylor), 159, 175.
- External gravitational and electromagnetic fields of rotating bodies (Clark), 488.
- Far ultra-violet absorption spectra of the hydrides and deuterides of sulphur, selenium and tellurium and of the methyl derivatives of hydrogen sulphide (Price, Teegan & Walsh), 600.
- Flashing, an investigation of the effect (Bremner), 305.
- Forces on a submerged spheroid moving in a circular path (Havelock), 297.
- Formation of a blast wave by a very intense explosion. I. Theoretical discussion. (Taylor), 159.
- Formation of a blast wave by a very intense explosion. II. The atomic explosion of 1945 (Taylor), 175.
- Frank, F. C. Radially symmetric phase growth controlled by diffusion, 586.
- Frank, F. C. & van der Merwe, J. H. One-dimensional dislocations. IV. Dynamics, 261.
- Friction and surface deformation of metals at light loads (Whitehead), 109.
- Gaydon, A. G. & Wolfhard, H. G. Spectroscopic studies of low-pressure flames. III. Effective rotational temperatures and excitation mechanism for C_2 bands, 561.
- Gaydon, A. G. & Wolfhard, H. G. Spectroscopic studies of low-pressure flames. IV. Measurements of light yield for C_2 bands, 570.
- Griffiths, J. H. E. *See* Bagguley & Griffiths.
- Grimley, T. B. The contact between a solid and an electrolyte, 40.
- Havelock, T. H. The forces on a submerged spheroid moving in a circular path, 297.
- Hinshelwood, Sir Cyril. *See* Stubbs & Hinshelwood; *also* Bardwell & Hinshelwood.
- Hooke, Robert. Wilkins Lecture (Andrade), 439.
- Horse met-myoglobin, crystal structure. I (Kendrew), 62.
- Inelastic scattering of protons by magnesium, aluminium and other light elements (Rhoderick), 348.
- Instability of liquid surfaces when accelerated in a direction perpendicular to their planes. I. (Taylor), 192.
- Integration of the laminar boundary layer equation. I. Motion of an elliptic cylinder-separation (Meksyn), 268.
- Integration of the laminar boundary layer equation. II. Retarded flow along a semi-infinite plane (Meksyn), 279.
- Interaction representation in the quantum theory of fields (de Wet), 284.
- Ionization produced by metallic salts in flames. I (Belcher & Sugden), 480.
- Jahn, H. A. Theoretical studies in nuclear structure. I. Enumeration and classification of the states arising from the filling of the nuclear d -shell, 516.
- Kemball, C. The adsorption of vapours on mercury. IV. Surface potentials and chemisorption, 377.
- Kendrew, J. C. The crystal structure of horse met-myoglobin. I. General features: the arrangement of the polypeptide chains, 62.
- Kinetics of the thermal decomposition of normal paraffin hydrocarbons. II. Comparative measurements on the series from propane to n -decane (Stubbs & Hinshelwood), 18.

- Laminar boundary layer equation, integration of. I and II (Meksyn), 268, 279.
- Liquid surfaces, instability of, when accelerated perpendicular to their planes (Taylor), 192.
- Magnetic field in a conducting liquid in turbulent motion (Batchelor), 405.
- Martyn, D. F. Cellular atmospheric waves in the ionosphere and troposphere, 216.
- Meksyn, D. Integration of the laminar boundary layer equation. I. Motion of an elliptic cylinder—separation, 268.
- Meksyn, D. Integration of the laminar boundary layer equation. II. Retarded flow along a semi-infinite plane, 279.
- Mercury, adsorption of vapours on (Kemball), 377.
- Merton, Sir Thomas. On the reproduction and ruling of diffraction gratings, 187.
- Metals, surface deformation and friction at light loads (Whitehead), 109.
- Met-myoglobin (horse) crystal structure. I (Kendrew), 62.
- Note on the problem of a rotating mass of perfect fluid in relativity mechanics (Clark), 510.
- Nuclear structure, theoretical studies, I (Jahn), 516.
- One-dimensional dislocations. IV. Dynamics (Frank & van der Merwe), 261.
- Oxidation of gaseous methyl ethyl ketone (Bardwell & Hinshelwood), 26.
- Paraffin hydrocarbons, kinetics of thermal decomposition. II. (Stubbs & Hinshelwood), 18.
- Paramagnetic resonance in copper sulphate (Bagguley & Griffiths), 366.
- Permeability of porous materials (Childs & Collis-George), 392.
- Perturbation theory, some theorems (Titchmarsh), 473.
- Podolanski, J. Unified field theory in six dimensions, 234.
- Polypeptide chains, arrangement (Kendrew), 62.
- Porous materials, permeability (Childs & Collis-George), 392.
- Presidential Address (Robinson), 1.
- Pressure pulse, dispersion of, in the atmosphere (Scorer), 137.
- Price, W. C., Teegan, J. P. & Walsh, A. D. The far ultra-violet absorption spectra of the hydrides and deuterides of sulphur, selenium and tellurium and of the methyl derivatives of hydrogen sulphide, 600.
- Protons, scattering of, by light elements (Rhoderick), 348.
- Quantum theory of fields, interaction representation (de Wet), 284.
- Radially symmetric phase growth controlled by diffusion (Frank), 586.
- Reproduction and ruling of diffraction gratings (Merton), 187.
- Rhoderick, E. H. The inelastic scattering of protons by magnesium, aluminium and other light elements, 348.
- Robinson, Sir Robert. Presidential Address, 1.
- Rotating bodies external gravitational and electromagnetic fields (Clark), 488.
- Rotating mass of perfect fluid in relativity mechanics (Clark), 510.
- Rotational temperatures and excitation mechanism for C_2 bands (Gaydon & Wolfhard), 561.
- Scorer, R. S. The dispersion of a pressure pulse in the atmosphere, 137.
- Six-dimensional unified field theory (Podolanski), 234.
- Slow oxidation of gaseous methyl ethyl ketone (Bardwell & Hinshelwood), 26.
- Some theorems on perturbation theory. II (Titchmarsh), 473.
- Specific energy of crystal boundaries in tin (Aust & Chalmers), 210.

- Spectroscopic studies of low-pressure flames. III. Effective rotational temperatures and excitation mechanism for C_2 bands (Gaydon & Wolfhard), 561.
- Spectroscopic studies of low-pressure flames. IV. Measurements of light yield for C_2 bands (Gaydon & Wolfhard), 570.
- Spheroid, submerged, the forces on, in a circular path (Havelock), 297.
- Spontaneous magnetic field in a conducting liquid in turbulent motion (Batchelor), 405.
- Stubbs, F. J. & Hinshelwood, Sir Cyril. The kinetics of the thermal decomposition of normal paraffin hydrocarbons. II. Comparative measurements of the series from propane to *n*-decane, 18.
- Studies on the ionization produced by metallic salts in flames. I. The determination of the collision frequency of electrons in coal-gas/air flames (Belcher & Sugden), 480.
- Sugden, T. M. *See* Belcher & Sugden.
- Supersonic flow past a body of revolution, far from the axis (Whitham), 89.
- Surface deformation and friction of metals at light loads (Whitehead), 109.
- Surface potentials and chemisorption (Kemball), 377.
- Taylor, Sir Geoffrey. The formation of a blast wave by a very intense explosion. I. Theoretical discussion, 159.
- Taylor, Sir Geoffrey. The formation of a blast wave by a very intense explosion. II. The atomic explosion of 1945, 175.
- Taylor, Sir Geoffrey. The instability of liquid surfaces when accelerated in a direction perpendicular to their planes. I, 192.
- Teegan, J. P. *See* Price, Teegan, & Walsh.
- Theoretical studies in nuclear structure. I. Enumeration and classification of the states arising from the filling of the nuclear *d*-shell (Jahn), 516.
- Thermal accommodation coefficients of gases. I. An investigation of the effect of flashing (Bremner), 305.
- Thermal accommodation coefficients of gases. II. Determinations at room and at liquid-oxygen temperatures (Bremner), 321.
- Tin, specific energy of crystal boundaries in (Aust & Chalmers), 210.
- Titchmarsh, E. C. Some theorems on perturbation theory. II, 473.
- Ultra-violet absorption spectra for the hydrides and deuterides of sulphur, selenium and tellurium and methyl derivatives of hydrogen sulphide (Price, Teegan & Walsh), 600.
- Unified field theory in six dimensions (Podolanski), 234.
- van der Merwe, J. H. *See* Frank & van der Merwe.
- Walsh, A. D. *See* Price, Teegan & Walsh.
- Whitehead, J. R. Surface deformation and friction of metals at light loads, 109.
- Whitham, G. B. The behaviour of supersonic flow past a body of revolution, far from the axis, 89.
- Whitrow, G. J. *See* Coxeter & Whitrow.
- Wilkins Lecture. Robert Hooke (Andrade), 439.
- Wolfhard, H. G. *See* Gaydon & Wolfhard.
- World-structure and non-Euclidean honeycombs (Coxeter & Whitrow), 417.

25 JAN. 1955

